Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A SHAP-Augmented Machine Learning Framework for Optimizing Iron-Carbon Micro-Electrolysis of PVA Wastewater

Xukui Wei
Xinjiang Institute of Technology
Subiy Akbar
Lanzhou University
Jinyuan Li
Xinjiang Institute of Technology
Xiao Zhang
Xinjiang Institute of Technology
Dongjie Chen
Xinjiang Institute of Technology
Shiqing Cao
Lanzhou Jiaotong University
Jian Zhou
zhoujian@lzjtu.edu.cn

Lanzhou Jiaotong University

Research Article

Keywords: Polyvinyl alcohol wastewater, Iron-carbon micro-electrolysis, Machine learning, Response Surface Methodology, Mechanistic interpretation

Posted Date: November 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11299105/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

While iron-carbon micro-electrolysis (ICME) is effective for treating polyvinyl alcohol (PVA) wastewater, its optimization remains notably constrained by two limitations of Response Surface Methodology (RSM): RSM's poor extrapolation capability and inability to reveal underlying governing mechanisms beyond local experimental domains. To break this bottleneck, this study establishes a triple-integrated paradigm that synergistically combines the experimental rigor of RSM, the predictive power of machine learning, and the mechanistic insights from SHAP interpretability analysis. A high-fidelity dataset of 1,423 samples was constructed via Box-Behnken design augmented with literature data. Among seven machine learning (ML)algorithms evaluated, XGBoost achieved superior predictive performance for COD removal rate(test set R^2 = 0.8045, RMSE = 6.06, MAPE = 11.25%), demonstrating robust generalization. Crucially, SHAP analysis decoded insights inaccessible to RSM: it identified the Fe/C ratio as the dominant structural factor governing system efficiency and delineated pH 2–3 as the optimal window for synergistic Fe/C ratio - pH interactions. The core innovation lies in the three-tiered methodological integration: RSM provides a rigorous experimental foundation; machine learning enables global, high-accuracy prediction; and SHAP transforms the "black-box" model into a discovery tool for extracting condition-dependent rules. This paradigm shifts the optimization strategy from local empirical tuning to system-level intelligent design, enabling a translatable framework for real-time control and the scale-up process of complex wastewater treatment.

1. Introduction

Polyvinyl alcohol (PVA) is a critical industrial raw material, but the high-concentration, bio-refractory wastewater generated during its production poses a severe challenge to conventional water treatment technologies [1, 2]. The characteristically high chemical oxygen demand (COD) and persistent organic compounds in PVA wastewater significantly hinder the efficacy of biological treatment processes [3, 4], making the development of efficient and cost-effective physicochemical pretreatment or advanced treatment technologies imperative. Iron-carbon micro-electrolysis (ICME) has emerged as a highly promising solution due to its low cost, operational simplicity, and effective treatment performance [5]. Its principle relies on the formation of numerous microscopic galvanic cells between iron scraps and carbon particles, and the synergy of electrochemical oxidation-reduction, flocculation, and adsorption not only degrades organic pollutants but also enhances wastewater biodegradability. However, the treatment efficiency of ICME is complexly governed by the interplay of multiple factors, including influent pH, Fe/C ratio, reaction time, and aeration rate. Currently, process optimization in this field heavily relies on traditional one-factor-at-a-time experiments and Response Surface Methodology (RSM) [6]. While RSM, through structured experimental designs like Box-Behnken, can effectively explore factor interactions and identify localized optimum conditions, its inherent limitations present significant bottlenecks for optimizing complex, nonlinear processes:

First, RSM relies on pre-defined (often quadratic) polynomial models, which may inadequately capture the highly nonlinear and non-monotonic behaviors commonly encountered in industrial applications [7].

Second, the predictive power of RSM is confined to the specific factor space defined by the experimental design, lacking reliable extrapolation capabilities and thus often failing under real-world fluctuations in influent quality or operational parameters.

Machine learning (ML) represents a paradigm shift from such traditional methodologies. ML algorithms (e.g., support vector machines, random forests) adaptively learn complex, nonlinear mappings directly from data, endowing models with superior predictive robustness and generalization abilities [8]. However, purely data-driven ML models often face challenges in interpretability and physical grounding due to their "black-box" nature and the potential lack of a rigorous experimental design foundation.

To simultaneously overcome the limitations of RSM and harness the power of ML, this study proposes a triple-integrated paradigm that synergistically combines: (1) the statistical rigor of RSM to generate a high-quality, reproducible experimental dataset; (2) the predictive power of advanced ML to transcend polynomial limitations and achieve globally robust predictions; and (3) interpretable AI techniques (SHAP) to transform the "black-box" model into a discovery tool for decoding underlying mechanisms. The core innovation of this work lies in leveraging SHAP to quantitatively resolve nonlinear parameter interactions — a capability absent in RSM. While RSM can identify statistically significant interactions (e.g., the BC interaction between aeration rate and reaction time, $F = 136.00$, $p < 0.01$), SHAP analysis reveals their conditional activation rules, thereby addressing RSM's core limitation of being unable to quantify contribution weights across the full operational domain.

Critically, this framework is designed to establish a complete, self-consistent evidence chain from macro-scale performance to molecular-level mechanisms. We integrate macro-kinetic analysis with time-resolved spectroscopic characterization (UV-Vis and FTIR) to validate the insights derived from ML, thereby bridging the gap between correlation and causation and enabling a mechanistic understanding of PVA degradation pathways.

Accordingly, the core objectives of this study are:

(1) To generate a comprehensive, statistically sound experimental foundation through rigorously designed RSM experiments.

(2) To develop and evaluate multiple ML models for accurately predicting the COD removal rate efficiency of the ICME process.

(3) To apply interpretable ML techniques (SHAP) to identify dominant parameters (e.g., Fe/C ratio) and quantify their nonlinear influences and interactions.

(4) To integrate UV-Vis and FTIR spectroscopic analyses to elucidate the molecular-level degradation pathway of PVA, providing physicochemical evidence that corroborates the data-driven insights.

(5) To systematically compare RSM and ML approaches, articulating their complementary roles and the transformative potential of their integration.

We anticipate that this paradigm will not only yield process parameters superior to those derived from traditional RSM but also enable a fundamental shift from local empirical tuning to global intelligent prediction and system-level design, thereby providing a robust, mechanism-guided framework for real-time control and reliable scale-up of advanced wastewater treatment technologies.

2. Materials and methods

2.1. RSM experimental section

2.1.1. Raw water quality

Synthetic wastewater was prepared to accurately simulate actual industrial effluent and ensure experimental reproducibility. The composition was based on the quality of filtered wastewater from a PVA production line of a chemical plant in Lanzhou, China. The synthetic wastewater was PVA-based. As summarized in Table 1, it exhibited high COD and low pH, characteristics representative of the challenges in treating refractory organic industrial wastewater. In addition to conventional parameters, total organic carbon (TOC) and total oxygen demand (TOD) were also measured to comprehensively characterize the organic content from multiple perspectives, thereby providing a more robust basis for evaluating treatment efficacy.

Table 1
Main water quality indicators of simulated PVA wastewater

Parameter	Value
PVA[a] Concentration	800mg/L
COD[b]	1350mg/L
TOD[c]	413mg/L
TIC[d]	22.7mg/L
TC[e]	435.7mg/L
pH	4.0 ± 1
[a] PVA Polyvinyl alcohol	
[b] COD Chemical Oxygen Demand	
[c] TOD Total Oxygen Demand	
[d] TIC Total Inorganic Carbon	
[e] TC Total Carbon	

2.1.2. Experimental reagents and iron-carbon micro-electrolysis materials

Table S1 details all chemical reagents and the primary reactive media, iron-carbon micro-electrolysis filler (ICMEF), used in the experimental system.

The core treatment material was a commercial ICMEF (model TMIE-1) obtained from Tianyi Environmental Materials Co., Ltd. (Ganzhou, China). As shown in Fig. 1, the spherical granules enable uniform flow distribution in the fixed-bed reactor, minimizing channeling. The mesoporous surface structure provides abundant Fe^0 corrosion sites, which drive PVA chain scission via micro-electrolysis. Its high porosity (≥ 55%) and appropriate specific surface area (≥ 1.2 m²/g) are conducive to an increased reaction interface and enhanced mass transfer efficiency, while its mechanical strength (barrel compressive strength ≥ 5 MPa) guarantees long-term stability under aeration. These properties collectively enhance the reaction interface, mass transfer efficiency, and long-term stability under aeration. Comprehensive physicochemical parameters are summarized in Table S2.

2.1.3. RSM design

We applied RSM to optimize PVA removal efficiency in the recalcitrant wastewater stream detailed in Section 2.1.1 (characterized by elevated COD and acidic pH). A three-factor Box-Behnken design (BBD) with three levels per factor and six center point replicates was selected over central composite design (CCD) because it reduces experimental workload while maintaining sufficient design points to support reliable quadratic model fitting. This design balanced statistical rigor with operational constraints [9]. Preliminary single-factor experiments identified pH, aeration rate, and reaction time as the key parameters exerting the greatest influence on the response (e.g., COD removal rate); these were therefore selected for further investigation. Other parameters, namely the ICMEF dosage (200 g) and the Fe/C ratio (3 : 1), were fixed at their empirically determined optimal levels. A three-factor, three-level Box-Behnken design was implemented with the center point set at pH 4.75, an aeration rate of 2.5 L/min, and a reaction time of 105 min. The factors and their corresponding levels for the response surface design are presented in Table 2.

Table 2
Factors and levels of response surface design experiments

Level	Factor A:pH	Factor B:Aeration rate(L/min)	Factor C:Reaction time(min)
-1	1	0	90
0	3.5	2.5	105
1	6	5	180

2.2. Machine learning

2.2.1. Dataset construction

The ML training data were curated from two sources: in-house experimental data from our laboratory and published, peer-reviewed studies on PVA wastewater treatment using ICME. After systematic collection, standardization, and quality filtering, a dataset comprising 1,423 valid samples was assembled. Although our experiments employed a specific commercial ICMEF, the literature data encompass a range of brands, models, and operating conditions within this material system. By integrating these diverse sources, we aim to train a model that captures common reaction behaviors across different setups, instead of overfitting to a single experimental protocol, thereby improving its generalizability.

Based on domain knowledge and a review of the literature, we selected the following operational parameters (pH, aeration rate, reaction time and Fe/C ratio) as input features for the model, as they are known to strongly influence treatment efficiency [10].The prediction target was the COD removal rate, a standard metric of overall treatment effectiveness. The selected input features and the rationale for their inclusion are summarized in Table 3.

Table 3
Definition of model feature variables and basis for selection

Feature variable	Variable type	Basis for selection
pH	Continuous	A critical factor that governs reaction kinetics and degradation pathways
Fe/C ratio	Continuous	A key material proportion determining the number and efficiency of microscopic galvanic cells
Reaction time	Continuous	Determines the contact duration between pollutants and reactive sites
Aeration rate	Continuous	Controls dissolved oxygen concentration and mass transfer efficiency, thereby modulating the rate and pathways of the micro-electrolysis reaction
COD removal rate	Continuous	Prediction target; a core indicator for comprehensively evaluating wastewater treatment effectiveness

2.2.2. Data cleaning and preprocessing pipeline

To ensure the accuracy and robustness of the ML models, we implemented a systematic preprocessing pipeline. First, any samples missing critical indicators required for modeling were excluded. For the remaining dataset, we assessed the extent of missing data for each feature: numerical features with a low proportion of missing values (< 5%) were imputed using the median to minimize the impact of outliers and preserve the central tendency of the distribution, while features with a high missing rate (> 50%) were removed entirely [11].Potential outliers for each parameter were initially identified using the Interquartile Range (IQR) method, with data points outside the range $(Q_1 - 1.5 \times IQR, Q_3 + 1.5 \times IQR)$ flagged as statistical outliers [12]. However, recognizing that such extremes may reflect valid but rare phenomena in the domain, we did not automatically remove all flagged points. Instead, a secondary review based on domain expertise was conducted to determine their validity. For instance, certain data points exhibiting an exceptionally high COD removal rate under specific pH and Fe/C ratio conditions, although statistically flagged as outliers, were deemed plausible based on the reaction context. Consequently, these observations were retained not as anomalies but as indicators of the process's high-efficiency regime under specific operational conditions. Boxplot analysis (Fig. S1) indicated that all parameters spanned the intended operational range, yet exhibited distinct distribution characteristics: reaction time displayed greater variability and included several high-value outliers, reflecting the process efficiency's sensitivity to operational conditions and its behavior under extreme scenarios. In contrast, the Fe/C ratio, pH, and aeration rate showed narrower distributions, suggesting tighter experimental control. The identified outliers, particularly those associated with plausible reaction conditions, were retained following a review informed by domain knowledge, ensuring the dataset preserves realistic system extremes necessary for ML models to learn complex nonlinear response patterns.

The original features in this study, reaction time, aeration rate, Fe/C ratio, and pH, have different units and numerical ranges. Directly using these raw values in a model can lead to slow convergence and undue influence from features with larger magnitudes. Consequently, Z-score standardization was applied to preprocess the data [13]. This technique removes the effect of differing units, ensures feature comparability, and helps satisfy the distributional assumptions of many ML algorithms[14].

Z-score standardization is defined as Eq. (1): $Z = \frac{x-\mu}{\sigma} \# (1)$

where x is the original feature value, μ is the mean, and σ is the standard deviation of the feature across all samples. After transformation, each standardized feature has a mean of 0 and a standard deviation of 1 [15].

As detailed in Table S3, after standardization, the mean values of the four features were all very close to 0 (on the order of 10^{-16} to 10^{-17}, effectively zero at computational precision), and their standard deviations were all very close to 1. This result is in perfect agreement with the theoretical expectation of Z-score standardization, confirming that the data preprocessing was performed correctly and effectively. The standardized data was subsequently used for ML modeling to ensure equitable treatment of all features, thereby enhancing both training efficiency and model generalizability.

A heatmap of the feature correlation matrix was generated to assess the effectiveness of the experimental design and the independence among the operational parameters (Fig. S2) [16]. All pairwise Pearson correlation coefficients had absolute values below 0.3, indicating only very weak linear relationships. This confirms that, under the employed Box-Behnken design, the four selected parameters exhibit negligible multicollinearity. Such independence enhances the reliability of subsequent ML model training, as each feature contributes unique information, thereby improving both model interpretability and performance stability.

2.2.3. Algorithm selection and model training

To establish a robust benchmark for predicting chemical oxygen demand (COD) removal efficiency, we evaluated a systematic sequence of machine learning models. This progression began with linear regression as a baseline for interpretability, followed by fundamental nonlinear models (decision tree, SVM, and KNN) to capture complex process relationships. We then implemented three ensemble strategies, including random forest (a bagging method), lightGBM, and XGBoost (both gradient boosting variants), to quantify the performance gains attributable to ensemble learning. Consequently, we assessed the following seven machine learning models, each selected for its distinct applicability to the ICME process:

- Linear regression was established as a performance benchmark due to its simplicity and high interpretability [16–18]. Any complex model was required to significantly outperform it to be considered valuable. Furthermore, its performance acts as a probe: poor results intuitively indicate strong nonlinear characteristics in the ICME process [19].

- Decision tree was selected as a fundamental, interpretable nonlinear model. It served as a baseline to quantify the performance gain achieved by subsequent, more complex tree-based ensemble methods (like random forest and gradient boosting trees), thus highlighting the value of ensemble strategies [20].
- SVM was employed because it is well-suited for problems with limited samples and can capture potential complex nonlinear relationships through kernel function mapping, exploring global optimal solutions beyond linear assumptions [21, 22].
- KNN, a "lazy learning" algorithm relying on local sample similarity, was introduced to investigate the local smoothness and predictability of the process efficiency [23]. Its performance indicates whether COD removal rate is predominantly governed by local operating conditions within the parameter space.
- Random forest, a representative bagging ensemble method, was chosen because it builds multiple decision trees and aggregates their results, effectively reducing variance, preventing overfitting, and handling high-dimensional features. This allowed an evaluation of collective decision-making efficacy on our dataset [24].
- LightGBM was selected as an efficient gradient boosting implementation known for fast training and low memory usage [25]. Its primary role was to represent the boosting strategy (sequential error correction) and enable a direct comparison with the bagging approach (random forest) on this dataset.
- XGBoost, another efficient gradient boosting variant, was included due to its integrated regularization that controls model complexity and often enhances generalization [26]. A meticulous comparison with LightGBM was conducted to identify the optimal Boosting algorithm variant for this specific application.

The model selection process followed a rigorous pipeline to ensure robustness: The complete dataset (1,423 samples) was first randomly split into a training set (80%) and an independent test set (20%) [27]. Model construction and initial evaluation were performed on the training set. The final optimal model was selected solely based on its performance on the held-out test set, ensuring an unbiased assessment of generalization ability [28].

To comprehensively evaluate model performance, this study adopted a multi-dimensional evaluation index system:

(1) Coefficient of determination (R^2)

R^2 quantifies the proportion of variance in the target variable that is explained by the model. Its value ranges from 0 to 1, with values closer to 1 indicating a superior model fit [29]. R^2 is calculated using Eq. (2):

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

(2) Root mean square error (RMSE)

RMSE measures the deviation between predicted and actual values. It is sensitive to outliers, with smaller values indicating higher prediction accuracy [30]. RMSE is calculated using Eq. (3):

$$\mathrm{RMSE}\sqrt{=\frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}\left(\mathrm{y_i}\text{-}\widehat{\mathrm{y}}_{\mathrm{i}}\right)^2}\#\left(3\right)$$

(3) Mean absolute error (MAE)

MAE reflects the average magnitude of the absolute errors between predicted and observed values. It is less sensitive to outliers than RMSE and offers an intuitive interpretation of prediction error in the original units of the target variable [31]. MAE is calculated with Eq. (4):

$$\mathrm{MAE}=\frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}\left|\mathrm{y_i}\text{-}\widehat{\mathrm{y}}_{\mathrm{i}}\right|\#\left(4\right)$$

(4) Mean squared error (MSE)

MSE is the square of the RMSE, which means it amplifies the impact of larger prediction errors. It is calculated using Eq. (5):

$$\mathrm{MSE}=\frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}\left(\mathrm{y_i}\text{-}\widehat{\mathrm{y}}_{\mathrm{i}}\right)^2\#\left(5\right)$$

(5)Mean absolute percentage error (MAPE)

MAPE offers a dimensionless measure of prediction accuracy, expressed as a percentage. This allows for meaningful comparisons between models or datasets with varying magnitudes [32]. MAPE is defined in Eq. (6):

$$\mathrm{MAPE}=\frac{100\backslash\%}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}\left|\frac{\mathrm{y_i}\text{-}\widehat{\mathrm{y}}_{\mathrm{i}}}{\mathrm{y_i}}\right|\#\left(6\right)$$

(6)Train-test performance gap (ΔR²)

ΔR² measures the difference between the R² values on the training and test sets, serving as an indicator of model overfitting. A smaller ΔR² indicates better generalization ability. It is calculated using Eq. (7):

$$\Delta\,\mathrm{R}^2=\left|\mathrm{R}_{\mathrm{train}}^2\text{-}\mathrm{R}_{\mathrm{test}}^2\right|\#(7)$$

3. Results and discussion

3.1. RSM optimization results and mechanism interpretation

3.1.1. RSM experimental design results

The experimental results from the Box-Behnken design are presented in Table 4. The data were analyzed using "Design-expert 13" statistical software, which generated a regression equation describing the relationship between the independent variables and the COD removal rate. The equation is given as Eq. (8):

Table 4
Box-behnken design experimental design results

No.	Factor A: pH	Factor B: Aeration rate	Factor C: Reaction time	COD removal rate
1	0	0	0	50.9
2	-1	0	-1	20.31
3	1	-1	0	4.26
4	0	-1	-1	0
5	0	-1	1	1.13
6	-1	1	0	47.10
7	1	1	0	30.01
8	1	0	1	36.13
9	0	0	0	53.14
10	-1	-1	0	9.42
11	0	1	1	55.85
12	1	0	-1	2.01
13	0	1	-1	0.21
14	0	0	0	50.1
15	-1	0	1	50.08
16	0	0	0	53.78
17	0	0	0	53.14

$$Y = -8.34A^2 - 21.18B^2 - 16.74C^2 + 13.63BC$$
$$-2.98AB + 1.09AC - 6.81A + 14.79B + 15.8C + 52.21\# (8)$$

In this equation, Y represents the predicted COD removal rate (%), while A, B, and C denote the coded values for pH, aeration rate (L/min), and reaction time (min), respectively.

3.1.2. Analysis of variance and significance test

Analysis of variance (ANOVA) was performed on the experimental results from Table 4 to evaluate the statistical significance of the quadratic polynomial model fitted to the data [31]. A higher F-value and a lower p-value indicate greater model significance, where $p < 0.01$ denotes high significance, $p < 0.05$ indicates significance, and $p > 0.05$ suggests no significant effect. The ANOVA results for the COD removal rate (response variable Y) are presented in Table S4.

Based on the F-values, the factors influencing COD removal rate (in decreasing order of significance) were reaction time (C, F = 1036.41), aeration rate (B, F = 979.76), and pH (A, F = 118.47) within the selected experimental range. The regression model was highly significant (F = 481.65, $p < 0.0001$), confirming its reliability for prediction. This is further supported by the high adjusted R^2 (0.9897) and predicted R^2 (0.9454) values reported in Table S4. The small difference between these metrics indicates excellent agreement between the model predictions and actual observations, underscoring the model's robustness. Moreover, an adequacy precision ratio of 33.331, far exceeding the recommended threshold of 4, confirms the model's strong predictive capability.

To validate the reliability of the established response surface model, a systematic analysis of its residuals was conducted. The results of this analysis are summarized in Fig. S3, residual analysis of the response surface model relies primarily on two key diagnostic plots [33]. Fig. S3(a) shows a scatter plot of predicted versus actual COD removal rates, with the horizontal axis representing actual values and the vertical axis predicted values. Each point is colored according to its COD removal rate (ranging from 0 to 55.85). The data points closely follow the diagonal line, and the high R^2 value (0.9897) indicates excellent agreement between model predictions and experimental results, confirming the high predictive accuracy of the established response surface model. Fig. S3(b) shows the relationship between the externally studentized residuals and the predicted values. The residuals are randomly scattered around the zero line, with no systematic trend or nonlinear pattern, indicating no significant model misspecification. Furthermore, the spread of the residuals remains relatively constant across the range of predicted values and shows no evident funnel-shaped pattern, supporting the assumption of homoscedasticity. Although a few points exhibit larger residuals, their distribution appears random and is likely attributable to experimental error rather than a deficiency in the model. In conclusion, by synthesizing the findings from the diagnostic analyses above, it can be asserted that the quadratic polynomial model developed in this study is well-defined and statistically robust [34]. Its predictive outcomes are reliable and can be confidently applied to subsequent process parameter optimization and mechanistic interpretation.

3.1.3. Analysis of two-factor interaction effects

Three-dimensional response surface plots and their corresponding contour plots are employed to assess the interaction effects between pairs of experimental factors on the COD removal rate [35]. In contour plots, near-circular contours around the optimum suggest weak interaction, whereas elliptical contours indicate strong interaction [35]. The gradient of the 3D surface reflects the sensitivity of the response: a shallow slope implies low sensitivity to the factor, while a steep gradient signifies high sensitivity. Color gradients in both plot types denote the magnitude of the COD removal rate, with warmer

(or higher-value) regions corresponding to better performance [37]. Using this framework, the interaction effects for factor pairs AB, AC, and BC are visualized in Fig. 2 and analyzed below.

As shown in Fig. ($2a_1$) and Fig. ($2a_2$), the response surface for the pH – aeration rate interaction exhibits a distinct maximum COD removal rate within the experimental domain. At low pH (with reaction time fixed at the central level), COD removal rate initially increases with aeration, peaks, then declines. This behavior is attributed to enhanced electrode potential under acidic conditions [38]. Beyond the optimal aeration rate, air bubble coverage on the filler surface reduces solid-liquid contact and Fe^{2+} dissolution, lowering micro-electrolysis efficiency. In contrast, under alkaline conditions, COD removal rate rises monotonically but modestly with aeration, primarily due to physical stripping of PVA (average rate: 1.4 $mg{\cdot}L^{-1}{\cdot}min^{-1}$), as the micro-electrolysis reaction itself is suppressed at high pH. The contours are more closely spaced along the aeration axis than the pH axis, indicating greater sensitivity to aeration rate — a finding corroborated by ANOVA (F-value for B = 320.43 vs. F-value for A = 67.98). The significant AB interaction ($F = 6.51$, $p < 0.05$) further confirms their coupled influence.

As shown in Fig. ($2b_1$) and Fig. ($2b_2$), the response surface for the pH – reaction time interaction is convex but exhibits a gentler slope compared to other factor pairs, indicating a weaker interaction. With aeration rate fixed at its central level, the COD removal rate consistently follows an inverted U-shape trend with increasing pH, regardless of reaction time. At low pH, extended reaction time initially enhances COD removal rate due to more complete PVA degradation. However, the rate of increase diminishes after 2 hours as the rising pH during the reaction progressively suppresses micro-electrolysis, eventually halting the process. In contrast, under alkaline conditions, COD removal rate increases only marginally with prolonged time, as physical stripping (a pH-independent mechanism) becomes the dominant removal pathway. The contour plot reveals denser spacing along the reaction time axis (Factor C) than the pH axis (Factor A), visually confirming that reaction time has a stronger influence on COD removal rate. This observation aligns with ANOVA results: the F-value for C (333.14) exceeds that for A (67.98), and the AC interaction term is not significant ($F = 0.86$, $p > 0.05$), further supporting the weak interaction between pH and reaction time.

As shown in Fig. ($2c_1$) and Fig. ($2c_2$), the aeration rate–reaction time (BC) interaction exhibits the strongest influence among all factor pairs. With pH fixed at its central level, the system behavior varies significantly depending on reaction time. At short reaction times, the COD removal rate initially increases with aeration but subsequently declines after reaching a peak. The initial rise is attributed to the synergistic effects of micro-electrolysis and physical stripping. However, the limited time restricts the full development of these processes, resulting in relatively low overall removal. The subsequent decline at high aeration rates is caused by bubble shielding of the filler surface, which reduces solid-liquid contact and hinders iron ion dissolution. In contrast, under extended reaction times, the COD removal rate rises sharply with aeration to a distinct maximum. Sufficient time allows the micro-electrolysis reaction to proceed fully and enhances stripping efficiency. Additionally, adequate aeration promotes the cathodic reaction, provides vigorous mixing to prevent channeling and passivation, and mitigates particle fouling

by pollutants. At low aeration rates, the COD removal rate initially increases with reaction time but later declines. This is likely due to the dominance of acidic iron dissolution over micro-electrolysis at prolonged exposure, which destabilizes the system. Conversely, at high aeration rates, removal increases rapidly at first but eventually plateaus as pH rises and pollutant concentrations decrease. The contour plot reveals similar line spacing along both aeration rate (Factor B) and reaction time (Factor C) near the response peak, indicating comparable individual contributions. This aligns with their ANOVA results (F-value for B = 320.43, F-value for C = 333.14). Crucially, the BC interaction term is highly significant (F = 136.00, $p < 0.01$), confirming the strong synergy between aeration rate and reaction time.

To elucidate the degradation pathway of PVA during ICME, we monitored the time-dependent changes in key water quality parameters under the conditions optimized by RSM (Fig. 3). The results demonstrated a significant reduction in PVA, COD, and TOC concentrations over time, with most degradation occurring in the initial reaction phase. The fastest COD removal rate (steepest slope) during this period aligns with the conclusion that reaction time is critical for early-stage efficiency enhancement. After 120 minutes, all curves plateaued, indicating that the reaction reached equilibrium and further time extension yielded minimal efficiency gains. This supports the selection of optimal reaction duration to minimize resource waste. Notably, variations in total inorganic carbon (TIC) provided critical insights into the reaction mechanism. The initial increase in TIC concentration suggests that PVA underwent both decomposition and partial mineralization during ICME, yielding CO_2 and carbonate ions (e.g., CO_3^{2-} or HCO_3^-). This highlights the role of carbon species transformation in confirming that PVA degradation via iron–carbon micro-electrolysis primarily involves chemical oxidation and mineralization, rather than mere physical adsorption or phase transfer. This finding transcends the traditional COD-focused evaluation system, offering a mechanistic understanding of ICME that integrates chemical oxidation and carbon transformation pathways. The kinetic chart identifies reaction time as a key temporal variable determining degradation extent, consistent with SHAP analysis in the ML model, which ranked it as the second most influential feature after the Fe/C ratio. This synergy demonstrates that data-driven models can extract meaningful insights from complex datasets while validating the "RSM-based experimental foundation and ML-driven breakthrough" paradigm: RSM provides interpretable mechanistic hypotheses, whereas ML enables global exploration of nonlinear interactions and threshold effects.

To confirm the proposed oxidative degradation pathway at the molecular level and gain mechanistic insights into the reaction process, the study utilized UV-Vis and FTIR spectroscopy to monitor structural changes during degradation. A key finding from Fig. S4 is that the wastewater absorbance exhibited a non-monotonic, complex change with treatment time, in the order: effluent after 90 min of ICME treatment > effluent after 120 min treatment > effluent after 60 min treatment > effluent after 30 min treatment > untreated raw water. Fig. S4 highlights a non-monotonic absorbance pattern in wastewater with treatment duration, peaking at 90 minutes of ICME and declining afterward in this sequence: 90-minute-treated effluent > 120-minute-treated effluent > 60-minute-treated effluent > 30-minute-treated effluent > raw water. The absorbance peaked at 90 minutes before declining, indicating a dynamic degradation mechanism. Initially (0–90 min), PVA macromolecules were rapidly cleaved into small-

molecule intermediates (e.g., aromatic or conjugated structures) that absorb UV light strongly. Subsequently (90–120 min), these intermediates were further oxidized and decomposed, leading to the absorbance decrease. This trend, consistent with the continuous rise in TIC (Fig. 3), confirms a chemical oxidation pathway involving sequential steps.

The "generation and transformation of intermediate products" inferred from the UV-Vis spectra was unequivocally confirmed by the FTIR analysis. As shown in Fig. S5 (FTIR spectra), the infrared spectra of the treated water samples changed significantly compared to the untreated sample. On one hand, the stretching vibration peaks characteristic of the PVA backbone, namely O-H (~ 3300 cm^{-1}) and C-H (~ 2900 cm^{-1}), weakened markedly, confirming the destruction of the macromolecular structure. Crucially, new absorption peaks appeared around ~ 1650 cm^{-1} and ~ 1400 cm^{-1}, which are assigned to the characteristic vibrations of carboxylate groups ($–COO^-$). This directly identifies carboxylic acids as one of the key intermediate products detected by UV-Vis spectroscopy, thereby confirming the oxidative degradation nature of PVA from the perspective of functional group transformation.The generation and transformation of intermediate products inferred from UV-Vis spectra were clearly confirmed by FTIR analysis (Fig. S5). The infrared spectra of treated samples showed significant changes compared to the untreated PVA: the characteristic stretching vibration bands of the PVA backbone (O–H at ~ 3300 cm^{-1} and C–H at ~ 2900 cm^{-1}) weakened markedly, indicating macromolecular degradation. Notably, new absorption bands at ~ 1650 cm^{-1} and ~ 1400 cm^{-1} were observed, attributed to carboxylate groups ($–COO^-$). These findings directly identify carboxylic acids as key intermediates detected by UV-Vis spectroscopy, thereby confirming the oxidative degradation pathway of PVA through functional group transformation.

In summary, the concentration curves, UV-Vis spectra, and FTIR spectra form a coherent evidence chain: the concentration curves reveal the macroscopic removal efficiency and mineralization trend; the UV-Vis spectra, through their unique temporal sequence changes, capture the dynamic "generation-consumption" pattern of intermediate products during degradation; and the FTIR spectra ultimately confirm the chemical nature of the oxidation products at the molecular structure level. Together, they demonstrate that the degradation of PVA by ICME is a complex chemical oxidation process involving macromolecular chain scission, generation of carboxylic acid intermediates, and ultimate mineralization.

In summary, the concentration profiles, UV-Vis spectra, and FTIR spectra collectively form a cohesive evidence chain: the concentration profiles reveal macroscopic removal efficiency and the mineralization process; UV-Vis spectra, through their distinct temporal evolution, capture the dynamic "generation and consumption" pattern of intermediates; and FTIR spectra directly identify the chemical nature of oxidation products at the molecular level. This integrated analysis demonstrates that PVA degradation via ICME is a complex chemical oxidation process involving macromolecular chain scission, carboxylic acid intermediate formation, and ultimate mineralization.X

3.2. Machine learning model performance comparison

The performance of several mainstream ML algorithms was systematically evaluated for predicting the COD removal rate using five metrics: R^2, RMSE, MAE, MSE, and MAPE. The corresponding results are summarized in Table 5. Among the evaluated models, the gradient boosting algorithm XGBoost achieved the best performance, attaining the highest R^2 (0.8045) and the lowest RMSE and MAE values, indicating its superior predictive accuracy. This superior performance can be attributed to XGBoost's ability to effectively capture complex nonlinear relationships and feature interactions while mitigating overfitting. Combined with interpretability analyses (SHAP values), the model offers insights into the key factors governing PVA degradation in the ICME process. Other ensemble methods, including LightGBM and Random Forest, also delivered strong results, though slightly inferior to XGBoost—likely owing to XGBoost's explicit regularization terms and second-order boosting optimization [39]. In contrast, simpler models exhibited notable limitations: the Decision Tree showed poor generalization, likely due to overfitting, while Linear Regression failed to capture the underlying patterns, as expected from its linear assumptions. The Support Vector Machine (SVM) also underperformed on this dataset. The very low R^2 value of the linear model (0.1480), in stark contrast to the high accuracy of ensemble methods, strongly suggests that the PVA degradation process is highly nonlinear and complex. These results support the study's central premise that the quadratic polynomial model underlying traditional RSM is insufficient to adequately represent the complexity of this process, whereas ML approaches such as XGBoost can better capture its nonlinear dynamics and deliver more accurate predictions. Consequently, XGBoost was adopted for all subsequent modeling and analysis.

Table 5
Machine learning model performance comparison

Model	R^2	RMSE	MAE	MSE	MAPE
XGBoost	0.8045	6.0604	3.7119	36.7281	11.2468%
LightGBM	0.7928	6.2392	4.1385	38.9271	9.7000%
Random forest	0.7575	6.7494	3.3143	45.5549	9.1900%
SVM	0.7516	6.8303	3.6274	46.6535	11.6057%
KNN	0.6865	7.6743	3.8165	58.8955	10.8857%
Decision tree	0.4271	10.3734	8.0242	107.6073	25.3450%
Linear regression	0.1480	12.6510	10.4455	160.0483	31.9436%

3.3. In-depth examination of the optimal machine learning model

3.3.1. Predictive performance analysis

Comparison of the seven models confirmed that XGBoost demonstrated superior performance in predicting the COD removal rate during PVA wastewater treatment via ICME. However, reliance on

primary metrics (R^2 and RMSE) alone is insufficient for a comprehensive assessment of model reliability and practicality. Diagnostic scatter plots (predicted vs. measured values) and residual distribution plots were therefore generated to evaluate the predictive accuracy, error characteristics, and consistency of the XGBoost model, thereby supporting subsequent mechanistic interpretation. The scatter plot of predicted versus measured COD removal rates (Fig. S6a) shows that the vast majority of data points cluster closely around the $y = x$ line, indicating strong agreement. The fitted linear regression ($y = 0.807x + 7.704$) has a slope near unity and a small intercept, suggesting minimal systematic bias. Moreover, the high coefficient of determination (R^2 = 0.8045) between predicted and observed values indicates that the model captures over 80% of the variability in the data, confirming its strong predictive performance. The residual analysis for the XGBoost model (Fig. S6b) indicates that the prediction errors satisfy key regression assumptions. The residuals (measured minus predicted values) are randomly scattered around zero with no apparent pattern and exhibit homoscedasticity, as evidenced by a mean close to zero (− 0.172) and a standard deviation of 6.058. These characteristics suggest minimal systematic bias and consistent error variance across the range of predictions. Critically, no discernible patterns, such as funnel-shaped or curved trends, are observed in the residuals across the prediction range (0%–80% COD removal rate), which is consistent with the regression assumptions of homoscedasticity and the absence of systematic error. These results indicate that the XGBoost model flexibly approximates the underlying input–output relationships and yields unbiased predictions. Given its high predictive accuracy (R^2 = 0.8045) and well-behaved residuals, the model serves as a reliable tool for predictive modeling and can guide further investigation into the factors governing PVA degradation. However, any mechanistic interpretation must be supported by domain knowledge and complementary interpretability analyses.

The XGBoost-derived feature importance ranking is presented in Fig. S7. The analysis indicates that the Fe/C ratio is the dominant predictor of COD removal rate efficiency, exhibiting an importance score significantly higher than those of reaction time, initial pH, and aeration rate. This analysis demonstrates that the initial Fe/C ratio is the primary determinant of COD removal rate efficiency, as it governs the fundamental reaction kinetics of the ICME process. These results further reveal a critical limitation of conventional RSM, which typically fixes the Fe/C ratio and exclusively optimizes operational parameters (e.g., reaction time, initial pH, aeration rate). Unlike RSM, the XGBoost analysis demonstrates that the Fe/C ratio governs COD removal rate efficiency more decisively than operational adjustments, as it defines the foundational reaction kinetics of the micro-electrolysis system.

3.3.2. Interpretability analysis

Building on the global feature importance ranking (Fig. S7), SHAP analysis was applied to the optimal XGBoost model to (1) quantify local contributions of operational parameters and (2) assess whether the inferred feature–response relationships align with established ICME kinetics.

As shown in Fig. 4, feature importance based on SHAP analysis identifies the Fe/C ratio as the dominant factor influencing COD degradation efficiency, followed by reaction time, aeration rate, and pH. This order starkly contrasts with the RSM-ANOVA ranking (reaction time > aeration rate > pH) reported in Section

3.1.2. The discrepancy arises because the Box-Behnken design used in RSM held the Fe/C ratio constant, thereby excluding it from the ANOVA significance assessment. Consequently, RSM could only rank the adjustable operational parameters within a pre-optimized structural framework.

Feature importance based on SHAP analysis (Fig. 4) further reveals the direction and nonlinearity of each feature's influence:

Fe/C ratio: High ratios (red dots) predominantly yield negative SHAP values, indicating an inhibitory effect, likely due to iron overload, which impedes effective galvanic cell formation and promotes filler hardening.

Reaction time: Longer durations (red dots) consistently associate with positive SHAP values, reflecting the multi-step nature of PVA degradation (oxidation, reduction, coagulation), which requires sufficient residence time.

Aeration rate: The scattered distribution suggests a nonlinear response, moderate to high aeration tends to enhance performance by improving oxygen supply and mixing, whereas excessive aeration may induce shear-induced disruption.

pH: Strongly acidic conditions (red dots) correspond to high positive SHAP values, while alkaline conditions (blue dots) uniformly suppress degradation, which is fully consistent with the acid-dependent activation of iron filings in micro-electrolysis systems.

These findings point to a fundamental distinction between operational and structural determinants of process performance. Operational parameters, such as reaction time, aeration rate, and pH, are optimizable within a fixed reactor configuration, and their relative influence within a constrained experimental window is reliably captured by RSM. In contrast, structural parameters like the iron-carbon ratio define the intrinsic capacity of the micro-electrolysis system; their global importance only becomes evident when the model is trained on data spanning a broader range of conditions, as enabled by the integration of experimental, historical, and literature-derived datasets. This explains why SHAP analysis identifies the iron-carbon ratio as the dominant factor, whereas RSM, by design, held it constant and thus could not assess its effect. Rather than indicating inconsistency, this divergence reflects the complementary perspectives offered by traditional design-of-experiments and data-driven modeling: the former excels at optimizing controllable settings under fixed system architecture, while the latter can uncover latent governing factors that lie outside the original experimental scope. The SHAP dependence patterns further support this interpretation, revealing nonlinear responses and threshold effects that align with known physicochemical principles of ICME.

Building on the global insights from the feature importance based on SHAP analysis (Fig. 4), which identified the Fe/C ratio as the dominant predictor and revealed pronounced nonlinear effects, we employed SHAP dependence plots to deconstruct the specific response patterns and, more critically, the interactive relationships among key process parameters. This analysis shifts the focus beyond individual

feature importance to elucidate how parameters jointly govern treatment efficiency through synergistic or conditional interactions. The dependence plots (Fig. 5) uncover four pivotal interaction mechanisms that underlie the feasible operational window observed in the experimental data:

The SHAP dependence plot (Fig. 5a) reveals a strong and condition-dependent interaction between aeration rate and reaction time that significantly influences system performance. Across the tested range of 2.3–8 L/min, aeration consistently enhances COD removal rate, with its contribution peaking at approximately 4 L/min. This optimum corresponds to the threshold needed to maintain adequate dissolved oxygen for cathodic reactions while minimizing bubble coverage that could block active sites on the iron–carbon filler. Importantly, the benefit of aeration is highly dependent on reaction time: the greatest positive SHAP contributions, indicating the strongest enhancement of predicted COD removal rate, are observed only when moderate aeration (approximately 4 L/min) is combined with extended residence times of 300 to 400 minutes, as shown by the clustering of light cyan-colored points near the peak. This suggests that short hydraulic retention times limit the effectiveness of aeration, whereas longer durations allow the micro-electrolysis process to proceed more completely, supported by concurrent gas–liquid mass transfer. At aeration rates above 4 L/min, SHAP values gradually decline, likely because high gas hold-up displaces the liquid phase and reduces contact between reactive media and pollutants. Nevertheless, no negative SHAP values appear even at 8 L/min, confirming that excessive aeration does not inhibit or destabilize the system. Collectively, these results define an effective operating window where moderate aeration and prolonged contact time jointly achieve high treatment efficiency—a regime identifiable through data-driven modeling but inaccessible to conventional experimental designs.

The SHAP dependence plot (Fig. 5b) reveals a nuanced relationship between reaction time and pH that governs the system's predictive response. Across the tested reaction time range of 80–600 min, the SHAP values associated with reaction time exhibit a non-linear pattern, transitioning from predominantly negative contributions at short times (< 80 min) to positive contributions at extended durations (> 150 min). This reversal highlights a critical threshold where reaction time shifts from suppressing to enhancing the predicted outcome. Notably, the magnitude and direction of this effect are strongly modulated by pH: while low pH conditions (1–4, blue-purple points) consistently associate with negative SHAP values at short times (< 80 min), they transition to moderate positive contributions at extended times (400–500 min). In contrast, high pH conditions (6–8, green points) demonstrate a divergent behavior, which is exhibiting extremely high positive SHAP values at short times but maintaining stable positive contributions even at long durations (600 min). This pH-dependent duality indicates that the system's response to reaction time varies depending on the underlying chemical mechanisms: low pH likely favors slow kinetics requiring prolonged exposure, whereas high pH enables rapid but transient reactions. The clustering of data points between 80–300 min indicates this is the operational window where the model's predictions are most sensitive to reaction time variations, likely reflecting the period during which key reactions (e.g., acid-base equilibria, redox transformations) reach their peak activity. Importantly, the absence of negative contributions at long times (> 400 min) across all pH ranges confirms that extended reaction durations do not destabilize the system, but rather enable sustained

performance through different mechanistic routes depending on pH. These findings define an actionable operational regime where pH acts as a switch to determine whether reaction time promotes or hinders the predicted outcome, a relationship that would be difficult to detect in conventional experiments.

The SHAP dependence plot for pH (Fig. 5c) shows a sharply defined optimal contribution window between pH 2–3, where SHAP values peak. This range aligns with the well-established optimal pH for sustained iron corrosion and efficient Fe^{2+} release in ICME systems, enabling subsequent Fenton-like oxidation of organic pollutants. The positive impact within this window is further regulated by the Fe/C ratio: the highest SHAP values coincide with moderate Fe/C ratios (2–4, purple-blue points), suggesting synergistic activation of redox processes between iron and carbon surfaces. In contrast, samples with high Fe/C ratios (4–7, green-yellow points) exhibit near-zero or negative contributions at low pH. This implies that a high iron load is detrimental under strongly acidic conditions but becomes less inhibitory at higher pH. Outside the narrow pH 2–3 range, SHAP values drop sharply, indicating a rapid loss of treatment efficiency: at pH < 2, excessive hydrogen evolution and inefficient Fenton chemistry limit the effective utilization of dissolved iron, while at pH > 4, the formation of passivating iron hydroxide layers suppresses further corrosion and blocks reactive sites. The high density of data points within the optimal pH range confirms its robustness. While the advantage of mildly acidic conditions is well documented, SHAP analysis quantifies the exact boundaries of this optimal window and demonstrates how its effectiveness is modulated by the Fe/C ratio, which is providing a multivariate refinement of traditional operating rules.

A narrow optimal Fe/C ratio is strongly modulated by seration rate (Fig. 5d): The SHAP dependence plot pinpoints a distinct optimum for treatment efficiency near an Fe/C ratio of 1, where SHAP values peak. This optimum aligns with the theoretical stoichiometry required for efficient electron transfer between iron and carbon in micro-electrolysis systems. The positive impact is critically dependent on aeration: the highest SHAP values are consistently observed at moderate aeration rates, indicating that optimal O_2 supply synergistically enhances redox cycling when iron and carbon act as complementary electrodes. Deviations from this balanced regime significantly reduce efficiency. At low aeration insufficient oxygen supply likely limits Fe^{2+} oxidation, diminishing treatment performance even at the optimal ratio. Conversely, high aeration rates are counterproductive, yielding negative SHAP values across all ratios, which is likely due to disruption of reactive micro-environments at the iron–carbon interface or reduced effective contact time. Similarly, Fe/C ratios outside the narrow window of 0.8–1.2 lead to sharply lower efficiency, reflecting either insufficient reactive sites or iron-induced passivation and electron pathway short-circuiting. The concentration of high-density data points around a ratio of 1 underscores the robustness of this operational zone.

Thus, high treatment efficiency is only achieved when both the Fe/C ratio and aeration rate are tuned to their respective optimal ranges, which constitute a coupled condition emerging from their nonlinear interaction and would be missed by single-factor experiments.

3.3.3. Robustness test

To evaluate the suitability and stability of the constructed XGBoost model for practical industrial environments, this section details a systematic robustness test. The robustness of a ML model refers to its ability to maintain stable performance under non-ideal conditions, such as data noise, feature missing, or parameter fluctuations. This capability is a key indicator for assessing the model's practical value and deployment potential [40].

This test specifically employed random feature missing as the perturbation method to simulate common issues like data transmission errors or partial information loss in real-world wastewater treatment processes. During testing, feature values in the test set were randomly set to zero at specific ratios, with a gradient of missing proportions from 0% to 50%. To eliminate the randomness of a single test, the experiment was independently repeated 10 times at each missing ratio level. Ultimately, model performance was characterized by the mean ± standard deviation of the R^2 scores from these 10 repeated experiments, yielding a stable performance degradation curve with error estimates.

The test results are shown in Fig. S8. The model's predictive performance (R^2) exhibited a monotonically decreasing trend as the missing proportion increased; however, the decay curve was relatively gradual. When the missing proportion was below 20%, the model's R^2 consistently remained above 0.4, with a coefficient of variation below 10%, demonstrating good stability. This threshold can be considered a "safety boundary" for the model in practical applications. Further analysis revealed that for every 10% increase in the missing proportion, the model's R^2 decreased by approximately 0.15 on average, without any precipitous performance drop. This indicates that the XGBoost model, through its sparse-aware algorithm for handling missing values, can effectively manage the impact of partial information loss. However, when the missing proportion exceeded 30%, the performance degradation accelerated (coefficient of variation > 9%), and the model's prediction reliability decreased significantly.

The above results indicate that the XGBoost model can maintain relatively robust predictive performance even when faced with varying degrees of data missing. This finding is highly significant for the model's deployment in real industrial settings, where data missing is often unavoidable. The stability demonstrated by the model with up to 20% missing data provides a reliable fault tolerance margin for its practical application.

3.4. Comparative analysis and discussion: RSM versus machine learning models

To systematically compare the applicability, advantages, and limitations of RSM and ML within the framework of this study, and to elucidate their inherently complementary relationship, a comparative analysis was conducted across multiple dimensions. The analysis reveals that RSM and ML constitute a synergistic optimization paradigm, not a substitutable one.

First, in structural factor identification, which reflects a shift from operational tuning to system design, RSM constrained by fixed structural parameters (Fe/C ratio = 3:1), identifies the BC interaction term (aeration rate × reaction time) as critical ($F = 136.00$, $p < 0.01$). This reflects RSM's implicit assumption of

unimodal optimization (e.g., single global optimum at pH = 3.5). In contrast, SHAP reveals the Fe/C ratio as the structural factor governing system performance ceiling, with its synergies (e.g., BC interaction) strictly dependent on this factor's optimal value. This represents a paradigm shift from "optimizing existing systems" (RSM) to "reconstructing system design" (SHAP).

Second, while RSM reveals statistically significant interactions, SHAP uncovers their conditional activation, shifting the perspective from global significance to contextual process rules. RSM confirms BC interaction existence (F = 136.00)c, while SHAP specifies its conditional activation: aeration benefit peaks only at reaction time > 300 min with moderate aeration (4 L/min). This specification of conditional activation by SHAP complements the RSM finding, offering a contextual control rule for process optimization.

Third, **while** RSM identifies a narrow pH window (3–4), SHAP refines this into a high-efficiency hypersurface: optimal pH (2–3) manifests only with moderate Fe/C ratio. Optimization shifts from "meeting single parameter thresholds" to "coordinating parameter interactions".

Fourth, RSM identifies a non-linear COD removal rate trend with aeration rate under short reaction times, peaking at intermediate rates before declining, but SHAP uncovers that this trend stems from reaction time's stage-dependent influence: contributions shift from negative (< 80 min, mass transfer limitation) to positive (> 150 min, reaction kinetics dominance). Thus, SHAP deciphers RSM's observation mechanistically, validating its hypotheses without model bias.

Finally, RSM's high predicted R^2 (0.9454) reflects excellent local interpolation within its design space (Box-Behnken design), while XGBoost's R^2 (0.8045) on heterogeneous test sets demonstrates robust global generalization. This epitomizes ML's core advantage: shifting from "locally optimal fitting" to "globally robust prediction."

In summary, RSM provides an indispensable foundation for mechanistic insight within predefined operational windows, while ML (via SHAP) reveals structural drivers, multi-modal operational domains, and conditional interaction rules. A comprehensive comparison of their fundamental differences across multiple dimensions is summarized in Table S5.Their synergy enables a paradigm where process optimization evolves from local empirical tuning to system-level intelligent design, directly addressing real-world wastewater variability.

4. Conclusions

This study establishes a synergistic framework integrating RSM with interpretable ML (XGBoost-SHAP) to advance the optimization of the ICME process for PVA wastewater treatment. Our analysis demonstrates that RSM and ML constitute complementary, not substitutable paradigms:

(1) From operational tuning to system design: SHAP analysis identifies structural factors (e.g., Fe/C ratio) as the ultimate governors of the system's performance ceiling, a perspective cannot capture due to

its fixed-design constraints (e.g., Fe/C ratio = 3:1).

(2) From static significance to contextual rules: SHAP-derived conditional activation rules (e.g., aeration efficacy peaking only at reaction times > 300 min) transform RSM's statistically significant interactions into actionable, context-aware process guidelines.

(3) SHAP refines RSM's narrow pH range (3–4) into a pH contribution domain (2–3) contingent on moderate Fe/C ratio, transforming optimization from "meeting single parameter thresholds" to "coordinating parameter interactions".

(4) From phenomenological trends to mechanistic thresholds: SHAP uncovers stage-dependent parameter effects (e.g., reaction time's transition from inhibitory to promotive beyond 80 min), revealing dynamic mechanistic thresholds that explain RSM-observed nonlinearities.

(5) From local interpolation to global generalization: The ML model demonstrates robust generalization capability (XGBoost R^2 = 0.8045 on heterogeneous data), complementing RSM's high-fidelity local interpolation (R^2 = 0.9454 within Box-Behnken design).

This framework closes the evidence chain from macroscopic kinetics (COD/TOC trends), through intermediate dynamics (UV-Vis evolution), to molecular mechanisms (FTIR confirmation of carboxylic acid formation), enabling a paradigm shift from "local empirical tuning" to "system-level intelligent design." It provides a robust, mechanism-guided methodology for real-world wastewater treatment, addressing variability through structural factor control, regulatory parameter optimization, and condition-aware operational rules, thereby advancing energy-efficient precision in water treatment beyond trial-and-error approaches.

Declarations

Funding

Not applicable.

Author Contribution

X.W. and J.Z. conceived the study and designed the experiments. X.W. performed the iron-carbon micro-electrolysis experiments and collected the data. S.A. and J.L. developed the machine learning models and conducted the SHAP analysis. D.C. and S.C. assisted with data validation and visualization. X.W. wrote the original draft of the manuscript. J.Z. reviewed, edited, and supervised the project. All authors read and approved the final manuscript.

Data Availability

All data generated or analvzed during this study are included in this published article and its supplementaryinformation files.Additional raw data are available from the correspondingauthor on reasonable request.

References

1. Jin, Z., et al., Construction of ZrO2-CeO2 composite UF membranes for effective PVA recovery from desizing wastewater. Separation and Purification Technology, 2023. 306: p. 122672. https://doi.org/10.1016/j.seppur.2022.122672
2. Piao, H., et al., Tubular nanofiber membranes combined with Z-scheme CuS@Co3S4 heterojunction catalyst for high-efficient removal of polyvinyl alcohol from waste water with high COD. Journal of Hazardous Materials, 2024. 480: p. 136354. https://doi.org/10.1016/j.jhazmat.2024.136354
3. Bian, L., Y. Dong and B. Jiang, Simplified creation of polyester fabric supported Fe-based MOFs by an industrialized dyeing process: Conditions optimization, photocatalytics activity and polyvinyl alcohol removal. Journal of Environmental Sciences, 2022. 116: p. 52–67. https://doi.org/10.1016/j.jes.2021.06.007
4. Sun, W., et al., Synergistic effect of ozonation and ionizing radiation for PVA decomposition. Journal of Environmental Sciences, 2015. 34: p. 63–67. https://doi.org/10.1016/j.jes.2015.01.020
5. Zhao, J., J. Gao and J. Liu, Preparation of a New Iron-Carbon-Loaded Constructed Wetland Substrate and Enhanced Phosphorus Removal Performance, in Materials. 2020. p. 4739. https://doi.org/10.3390/ma13214739
6. Wang, H., et al., Succinic acid production from xylose mother liquor by recombinant Escherichia coli strain. Biotechnology & Biotechnological Equipment, 2014. 28(6): p. 1042–1049. https://dx.doi.org/10.1080/13102818.2014.952501
7. Xie, L. and J. Yang, Performance Modulation of S-CO2 Brayton Cycle for Marine Low-Speed Diesel Engine Flue Gas Waste Heat Recovery Based on MOGA, in Entropy. 2022. p. 1544. https://doi.org/10.3390/e24111544
8. Wong, J., et al., Applying Machine Learning in Distributed Data Networks for Pharmacoepidemiologic and Pharmacovigilance Studies: Opportunities, Challenges, and Considerations. Drug Safety, 2022. 45(5): p. 493–510. https://dx.doi.org/10.1007/s40264-022-01158-3
9. Ezeogu, N., et al., Modelling Caffeine and Paracetamol Removal from Synthetic Wastewater Using Nanofiltration Membranes: A Comparative Study of Artificial Neural Networks and Response Surface Methodology, in Membranes. 2025. p. 222. https://dx.doi.org/10.1038/s41598-025-93775-y
10. Bystrzejewski, M., et al., Synthesis of carbon-encapsulated iron nanoparticles by pyrolysis of iron citrate and poly(vinyl alcohol): a critical evaluation of yield and selectivity. Nanotechnology, 2011. 22(31): p. 315606. https://doi.org/10.3390/membranes15080222
11. Wang, X., et al., Enhancing effluent biochemical oxygen demand prediction in wastewater treatment plants: An ensemble model of gradient boosting and LSTM. Journal of Environmental Chemical

Engineering, 2025. 13(5): p. 117897. https://dx.doi.org/10.1088/0957-4484/22/31/315606

12. Zhou, X., X. Wang and R. Guo, Assessment model of ozone pollution based on SHAP-IPSO-CNN and its application. Scientific Reports, 2025. 15(1): p. 3404. https://dx.doi.org/10.1038/s41598-025-87702-4
13. Joo, C., et al., Machine Learning Approach to Predict Physical Properties of Polypropylene Composites: Application of MLR, DNN, and Random Forest to Industrial Data, in Polymers. 2022. p. 3500. https://doi.org/10.3390/polym14173500
14. Dargi, M., E. Khamehchi and J. Mahdavi Kalatehno, Optimizing acidizing design and effectiveness assessment with machine learning for predicting post-acidizing permeability. Scientific Reports, 2023. 13(1): p. 11851. https://dx.doi.org/10.1038/s41598-023-39156-9
15. Guohao, Z., et al., Optimization of convolutional neural networks for predicting water pollutants using spectral data in the middle and lower reaches of the Yangtze River Basin, China. Journal of Mountain Science, 2025. 22(08): p. 2851–2869. https://dx.doi.org/10.1007/s11629-024-9175-x
16. Maamoun, I., et al., Insights into machine-learning modeling for Cr(VI) removal from contaminated water using nano-nickel hydroxide. Separation and Purification Technology, 2023. 308: p. 122863. https://doi.org/10.1016/j.seppur.2022.122863
17. Shaharoona, B., et al., The role of urbanization in soil and groundwater contamination by heavy metals and pathogenic bacteria: A case study from Oman. Heliyon, 2019. 5(5). https://dx.doi.org/10.1016/j.heliyon.2019.e01771
18. Wang, J., et al., Prediction of CO2 solubility in deep eutectic solvents using random forest model based on COSMO-RS-derived descriptors. Green Chemical Engineering, 2021. 2(4): p. 431–440. https://doi.org/10.1016/j.gce.2021.08.002
19. Qaiser, F.U.R., et al., Characterization and health risk assessment of arsenic in natural waters of the Indus River Basin, Pakistan. Science of The Total Environment, 2023. 857: p. 159408. https://doi.org/10.1016/j.scitotenv.2022.159408
20. Lv, J., et al., Augmented machine learning for sewage quality assessment with limited data. Environmental Science and Ecotechnology, 2025. 23: p. 100512. https://doi.org/10.1016/j.ese.2024.100512
21. Da, T., et al., Prediction of uranium adsorption capacity on biochar by machine learning methods. Journal of Environmental Chemical Engineering, 2022. 10(5): p. 108449. https://doi.org/10.1016/j.ese.2024.100512
22. Wijakmatee, T. and H. Matsumoto, Physics-informed COSMO-SAC-machine learning framework for predicting ammonia solubility in choline chloride-based deep eutectic solvents. Separation and Purification Technology, 2026. 381: p. 135561. https://doi.org/10.1016/j.seppur.2025.135561
23. Shamshad, N., et al., Advanced KNN-based cost-efficient algorithm for precision localization and energy optimization in dynamic underwater sensor networks. Scientific Reports, 2025. 15(1): p. 2182. https://dx.doi.org/10.1038/s41598-025-86266-7

24. Venkatraman, V. and B.K. Alsberg, Predicting CO2 capture of ionic liquids using machine learning. Journal of CO2 Utilization, 2017. 21: p. 162–168.. https://doi.org/10.1016/j.jcou.2017.06.012
25. Wang, Q., et al., Advancing LightGBM with data augmentation for predicting the residual strength of corroded pipelines. npj Materials Degradation, 2025. 9(1): p. 128. https://dx.doi.org/10.1038/s41529-025-00673-9
26. Liu, H., et al., A generic machine learning model for CO2 equilibrium solubility into blended amine solutions. Separation and Purification Technology, 2024. 334: p. 126100. https://doi.org/10.1016/j.seppur.2023.126100
27. Zhao, C., et al., Reveal the main factors and adsorption behavior influencing the adsorption of pollutants on natural mineral adsorbents: Based on machine learning modeling and DFT calculation. Separation and Purification Technology, 2024. 331: p. 125706. https://doi.org/10.1016/j.seppur.2023.125706
28. Zhu, X., et al., Machine learning for the selection of carbon-based materials for tetracycline and sulfamethoxazole adsorption. Chemical Engineering Journal, 2021. 406: p. 126782. https://doi.org/10.1016/j.cej.2020.126782
29. Cao, S., et al., Machine learning-optimized interpretability analysis for identifying key drivers of NO3 lifetime variability. Journal of Environmental Sciences, 2025. https://doi.org/10.1016/j.jes.2025.06.064
30. Liu, C., et al., Prediction of heavy metal spatial distribution in soils of typical industrial zones utilizing 3D convolutional neural networks. Scientific Reports, 2025. 15(1): p. 396. https://dx.doi.org/10.1038/s41598-024-84545-3
31. Kurban, H., et al., Accelerating density of states prediction in Zn-doped MgO nanoparticles via kernel-optimized weighted k-NN. Scientific Reports, 2025. 15(1): p. 30524. https://dx.doi.org/10.1038/s41598-025-07887-6
32. Li, H., et al., An interpretable deep learning model for predicting the influent of wastewater treatment plants. Journal of Environmental Chemical Engineering, 2025. 13(4): p. 117437. https://doi.org/10.1016/j.jece.2025.117437
33. Tetteh, E.K., S. Rathilal and D.B. Naidoo, Photocatalytic degradation of oily waste and phenol from a local South Africa oil refinery wastewater using response methodology. Scientific Reports, 2020. 10(1): p. 8850. https://dx.doi.org/10.1038/s41598-020-65480-5
34. Pineda-Reyes, A.M., et al., Implementation of the emulsification-diffusion method by solvent displacement for polystyrene nanoparticles prepared from recycled material. RSC Advances, 2021. 11(4): p. 2226–2234. https://dx.doi.org/10.1039/D0RA07749F
35. Karichappan, T., S. Venkatachalam and P.M. Jeganathan, Optimization of electrocoagulation process to treat grey wastewater in batch mode using response surface methodology. Journal of Environmental Health Science and Engineering, 2014. 12(1): p. 29. https://dx.doi.org/10.1186/2052-336X-12-29

36. Li, S., et al., Preparation and characterization of superior hydrophilic PVDF/DA membranes by the self-polymerization approach of dopamine. Frontiers in Chemistry, 2023. Volume 11–2023. https://doi.org/10.3389/fchem.2023.1162348
37. Su, C., et al., Treatment of Cutting Fluid Waste using Activated Carbon Fiber Supported Nanometer Iron as a Heterogeneous Fenton Catalyst. Scientific Reports, 2018. 8(1): p. 10650. https://dx.doi.org/10.1038/s41598-018-29014-4
38. Lai, B., et al., Degradation of 3,3'-iminobis-propanenitrile in aqueous solution by Fe0/GAC micro-electrolysis system. Chemosphere, 2013. 90(4): p. 1470–1477. https://doi.org/10.1016/j.chemosphere.2012.09.040
39. Mishra, A., et al., Supervised Machine Learning and Physics Machine Learning approach for prediction of peak temperature distribution in Additive Friction Stir Deposition of Aluminium Alloy. PLOS ONE, 2025. 20(4): p. e0309751. https://doi.org/10.1371/journal.pone.0309751
40. Shen, R., et al., Robust prediction for characteristics of digestion products in an industrial-scale biogas project via typical non-time series and time-series machine learning algorithms. Chemical Engineering Journal, 2024. 498: p. 155582. https://doi.org/10.1016/j.cej.2024.155582

Figures

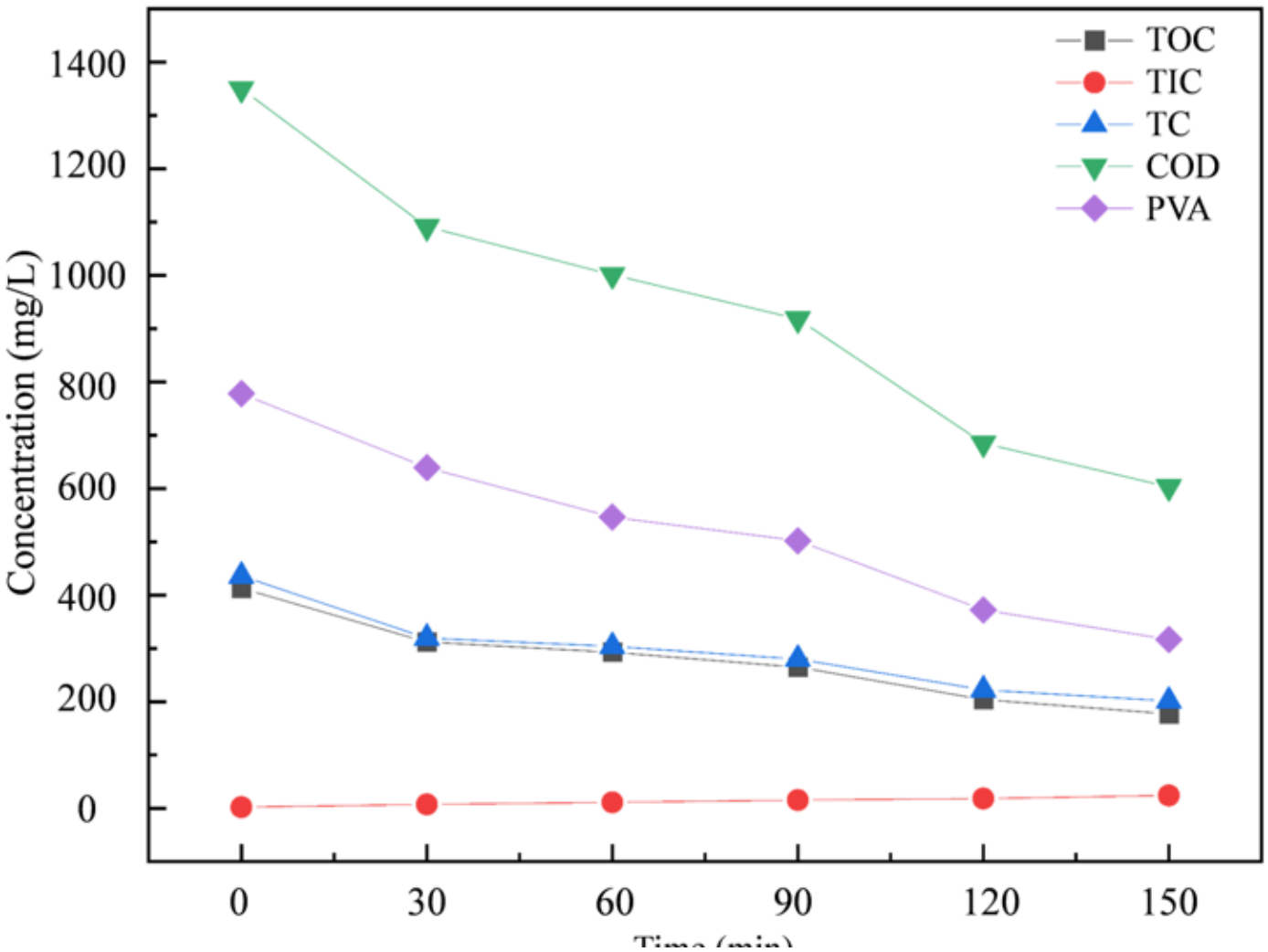

Figure 3

Temporal variations of organic and inorganic carbon components as well as pollutant indices during the treatment of PVA wastewater by iron-carbon micro-electrolysis.

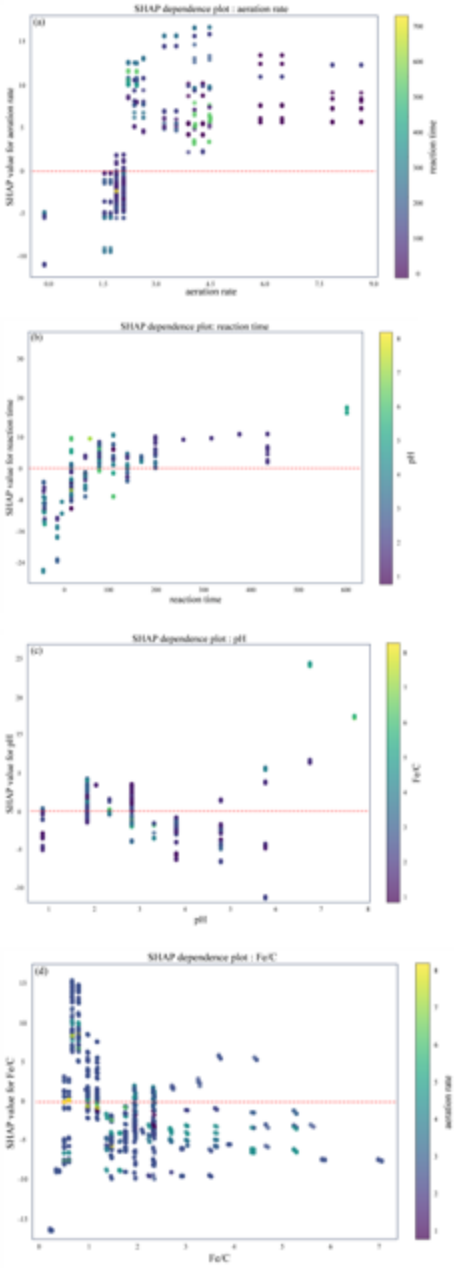

Figure 5

SHAP dependence plots for the key environmental factors in the ICME process: (a) iron-to-carbon ratio, (b) aeration rate, (c) initial pH, and (d) reaction time.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Tables.docx
- Supplementarymaterial.docx