THE PUBLISHING HOUSE
OF THE ROMANIAN ACADEMY

PROCEEDINGS OF THE ROMANIAN ACADEMY, Series A
Volume 27, Number 1/2026, pp. 101–112

DOI:

BATNet: A WAVELET-ENHANCED BiLSTM-TRANSFORMER WITH PRIOR KNOWLEDGE FOR WATER TREATMENT DOSING PREDICTION

Ting LIU[1], Jinhua LIU[1], Zhengshan DONG[2]

[1] Hunan College of Information, 410200, China

[2] School of Computer and Data Science, Minjiang University,350108, China

First author: Ting Liu, E-mail: 2020031@mail.hniu.cn

*Corresponding authors: Jinhua Liu, E-mail: 2020018@mail.hniu.cn and Zhengshan Dong, E-mail: dongzgs@mju.edu.cn

Abstract. Accurate coagulant dosing prediction in water treatment plants is challenging due to nonlinear, multivariable, and time-varying water-quality dynamics, while purely mechanistic or purely data-driven models often suffer from limited adaptability or interpretability. To address this, this study proposes BATNet(BiLSTM–Attention–Transformer Network with Wavelet Enhancement), a prior knowledge-driven dosing prediction framework that integrates an expert mechanistic model with data-driven residual correction. Specifically, wavelet transform is used to extract multiscale features from fluctuating inputs, a BiLSTM(Bidirectional Long Short-Term Memory)–Transformer backbone captures both short-term variations and long-range dependencies, and a weighted attention mechanism embeds expert-guided feature weighting to improve interpretability and robustness. Evaluations on two real-world datasets (Shanghai Xinghuo Water and Suzhou China-France Water) demonstrate consistent improvements over mainstream baselines (XGBoost, BP, LSTM, and BiLSTM), achieving test performance of $R^2 = 0.976$/MAE$= 0.078$/RMSE$= 0.108$ and $R^2 = 0.971$/MAE$= 0.088$/RMSE$=$ 0.124, respectively. These results indicate that BATNet provides accurate and reliable dosing prediction for practical water treatment operations.

Keywords: dosing prediction, prior knowledge model, wavelet transform, BATNet.
Mathematics Subject Classification: 68T07, 62M10, 93C10, 62P12.

1. INTRODUCTION

Traditionally, water treatment plants (WTPs) rely on empirically grounded physicochemical processes [1], where coagulant dosage is controlled by standardized equations and heuristics across coagulation, sedimentation, and filtration [2]. While crucial, coagulant dosage control is often subjective [3], relying heavily on operator experience and jar tests, leading to delayed responses and poor adaptability to rapid water-quality fluctuations [4]. Over-dosing can degrade treatment efficiency and increase operating costs [5]. Similarly, under-dosing may lead to insufficient pollutant removal and unstable treatment performance [6].

The increasing impact of climate change and extreme events further stresses WTP operations by inducing rapid turbidity surges and contaminant shocks [7]. While mechanistic models provide useful guidance, the nonlinear and multivariable nature of coagulation dynamics often makes simplified models inadequate under highly variable conditions [8].

Despite advancements in ICT and smart water initiatives, WTPs still largely rely on SCADA systems for supervision rather than predictive decision support [15]. Coagulation strategies generally follow feedforward or feedback control, but feedback control is reactive and may be insufficient during rapid changes in water quality [12]. Thus, reliable feedforward prediction of coagulant dosage is essential for improving operational robustness and treatment efficiency [13].

Data-driven ML approach have thus attracted increasing attention [14], as they can learn nonlinear mappings from operational data to dosing requirements without explicitly modeling reaction mechanisms [15]. Prior studies have applied BP neural networks, random forests (RF) [16], and support vector regression (SVR) to predict dosage or effluent quality [17]. Nevertheless, many models are trained on limited or low-frequency datasets and are often static, which restricts their practical deployment in water systems [18]. In addition, existing prediction models may fail to adequately capture seasonal variability and time-dependent dosing behavior, which further limits generalization and prediction accuracy [19].

To better leverage temporal dependencies in high-frequency streams, sequence models (e.g., RNNs/LSTMs [20] and Transformers) have been adopted for dosage prediction [21]. LSTM is effective for long-term dependencies and stable training, making it practical when data are limited [22]. However, challenges remain: single-architecture predictors may be insufficient for complex nonlinear coupling, and attention-based frameworks can outperform standalone LSTM models [23]; long monitoring sequences may contain redundancy and noise that obscure informative patterns [24]; and engineering deployment requires transparency [25], motivating explainable tools such as SHAP for feature attribution and decision support [26]. These issues motivate robust, interpretable, and temporally-aware models that generalize across diverse operating conditions [21].

To address the challenges of water treatment dosing, this study proposes BATNet(BiLSTM–Attention–Transformer Network with Wavelet Enhancement), a bidirectional LSTM-Transformer algorithm integrating expert knowledge with deep learning. Its key innovations lie in three aspects: **(1) Expert Knowledge Integration:** Mechanistic priors and operational expertise are embedded into the model, ensuring interpretability and alignment with practical dosing principles while capturing underlying water quality dynamics. **(2) Hybrid Model Architecture:** Wavelet Transform extracts short-term fluctuations, BiLSTM captures local temporal dependencies, and the Transformer models long-term context and feature interactions. Weighted Attention further incorporates expert-guided feature prioritization, improving accuracy under nonlinear, dynamic conditions. **(3) System-Level Dynamic Optimization:** Mechanistic predictions, wavelet features, and residual corrections are fused through a weighted, real-time optimization strategy, enabling robust, adaptive, and interpretable dosing predictions across varying water quality scenarios.

The remainder of this paper is organized as follows. Section 2 introduces the integration of prior expert models. Section 3 presents the proposed BATNet framework, including the problem formulation and model architecture. Section 4 describes the experimental settings and evaluation results on real-world water treatment datasets. Finally, Section 5 concludes the paper and discusses potential directions for future research.

2. Integration of Prior Expert Models

2.1 Data Acquisition

The dataset used in this study was collected from two water treatment plants in China, i.e., *Shanghai Xinghuo Water* and *Suzhou China-France Water*. It contains 180,624 operational records originally accumulated for expert-driven mechanistic modeling and coagulant dosing optimization to ensure compliance with water-quality standards. Key variables include inflow turbidity, inflow chemical oxygen demand (COD), and inflow temperature, which are widely recognized as critical indicators for alum dosing control: turbidity reflects suspended solids and directly affects dosage demand, COD indicates the organic pollution load, and temperature influences alum dissolution and reaction kinetics, thereby impacting coagulation efficiency. These variables, together with the corresponding dosing amounts, are continuously monitored via the plants' supervisory control and data acquisition (SCADA) systems, stored locally, and then transferred to centralized databases for management and analysis. Measurements are recorded at minute-level resolution and aggregated into 15-minute and hourly intervals for operational use; due to storage constraints, minute-level data are retained only for a limited period, whereas hourly data are archived permanently. For this study, 26,944 hourly samples were extracted for subsequent modeling and analysis.

2.2 Data Preprocessing and Analysis

To ensure reliable input-output relationships, raw operational data were preprocessed through data lagging, outlier/missing-value handling, and normalization. Data lagging aligned inflow measurements with downstream outcomes by considering the hydraulic retention time (HRT), calculated as the ratio of tank/pipeline volume to average flow rate. Outliers were screened using domain knowledge, and all records were retained due to the absence of physically unrealistic values. Missing values were imputed using a moving average (window size = 3) to maintain temporal continuity and reduce noise.

All variables were normalized to the range [0, 1] using min-max scaling:

$$x_s = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}, \quad x_s \in [0,1] \tag{1}$$

where x_i is the original value, and $x_{\min}$ and $x_{\max}$ are the minimum and maximum values in the dataset. Normalization ensures numerical stability and comparability across features.

To explore relationships among key variables, Pearson correlation analysis was performed:

$$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2}} \tag{2}$$

where X_i and Y_i are individual samples, and $\bar{X}$ and $\bar{Y}$ are their means.

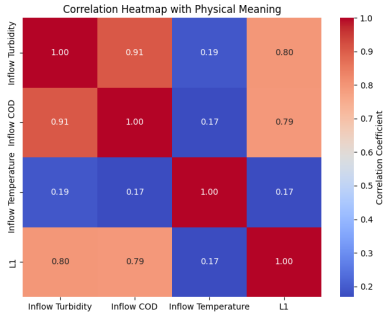


Fig. 1 – Correlation analysis among inflow variables and dosing output.

Fig. 1 shows the standardized correlations among inflow turbidity, COD, temperature, and alum dosing ($L1$). A strong positive correlation (0.90) is observed between turbidity and COD, suggesting suspended solids contribute to organic pollution. The correlation between temperature and dosing is weaker (0.55), primarily affecting dosing through reaction kinetics. Alum dosing shows moderate to strong correlations with both turbidity (0.77) and COD (0.85), highlighting their key roles in determining dosing requirements. These findings emphasize the need for adaptive dosing strategies to capture nonlinear interactions among water quality indicators.

2.3 Expert Mechanistic Model

Traditional alum dosing systems rely heavily on operator experience and expert judgment, with manual adjustments based on instrument readings and water quality indicators. While effective under stable conditions, this approach suffers from subjectivity, reliance on personal experience, and limited responsiveness to rapid or complex water quality changes.

An empirical mechanistic model was developed to describe alum dosing behavior, combining expert knowledge with real-time monitoring of turbidity, COD, temperature, and flow-related factors. The dosing process is formulated as:

$$Y_1 = 2.198 + \left(\frac{a + b \cdot Ef + a \cdot e^{-0.1 \cdot (A+x)}}{1 + e^{-0.1 \cdot (A+x)}} \right) \cdot Ef \cdot B + 0.007 \cdot C^2 - 0.35 \cdot C + \Delta \tag{3}$$

where Y_1 is the predicted alum dosage, A is raw-water turbidity, B is COD, C is temperature, and Ef is dosing efficiency.

The model uses monitored variables: turbidity (1–100 NTU), COD (1.5–6 mg/L), and temperature (5°C–35°C), with flow-related conditions influencing dosing. Empirical parameters a, b, x, and Δ control response sensitivity and adjust for deviations, with values tuned using historical data and real-time feedback.

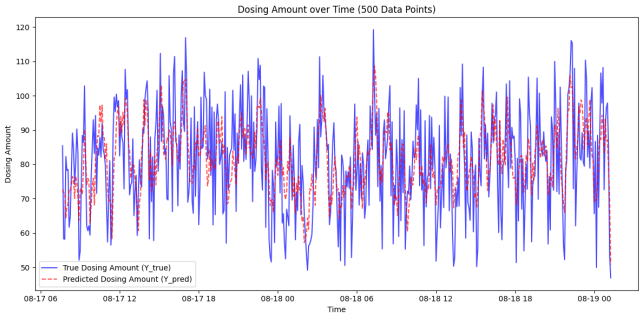


Fig. 2 – Mechanistic Model Dosing vs. Actual Dosing

As shown in Fig. 2 , the model's predictions follow the general trend of the true dosage, indicating it captures the main drivers under standard conditions. However, discrepancies occur during peaks and troughs, where rapid changes in water quality or unmodeled disturbances lead to errors. The model also responds more smoothly during high-demand periods, showing limited sensitivity to short-term dynamics.

Seasonal error patterns suggest time-varying effects that are difficult to capture with static parameters. While the expert mechanistic model provides valuable insight, its adaptability and precision are limited under dynamic conditions, highlighting the need for data-driven learning to improve robustness and accuracy.

3. Proposed Model

3.1 Problem Formulation

Let $\mathbf{x}_t \in \mathbb{R}^d$ denote the monitoring vector at time step t, where d is the number of input variables. Given a historical window of length L, the input sequence is defined as:

$$\mathbf{X}_t = [\mathbf{x}_{t-L+1}, \mathbf{x}_{t-L+2}, \ldots, \mathbf{x}_t]^\top \in \mathbb{R}^{L\times d}. \tag{4}$$

The goal is to predict the true alum dosage $y_t^{\text{true}} \in \mathbb{R}$. Instead of directly estimating the dosage, the proposed model adopts a prior-guided residual learning strategy. An expert model first provides an initial estimate:

$$y_t^{\text{cal}} = f_{\text{exp}}(\mathbf{x}_t), \tag{5}$$

and the residual is defined as:

$$r_t = y_t^{\text{true}} - y_t^{\text{cal}}. \tag{6}$$

To capture multiscale temporal variations, the input sequence is transformed by wavelet analysis:

$$\widetilde{\mathbf{X}}_t = \mathscr{W}(\mathbf{X}_t). \tag{7}$$

BATNet then learns the residual from the transformed sequence:

$$\hat{r}_t = f_\theta(\widetilde{\mathbf{X}}_t), \tag{8}$$

where $f_\theta(\cdot)$ denotes the proposed network with parameters θ.

The final prediction is obtained by

$$\hat{y}_t = y_t^{\text{cal}} + \hat{r}_t. \tag{9}$$

For clarity, scalars are denoted by italic letters, vectors by bold lowercase letters, and matrices by bold uppercase letters. This formulation is compatible with sliding-window prediction in real-world water-treatment systems.

3.2 Overview of the Model Architecture

Effective dosing prediction in water treatment systems is challenging due to the dynamic and nonlinear nature of water quality parameters. Traditional mechanistic models rely on expert knowledge but struggle with real-time variations, while purely data-driven models excel at pattern recognition but lack interpretability.

To address these challenges, this study proposes BATNet, a hybrid model combining expert priors with wavelet-enhanced feature extraction and a Bidirectional LSTM-Transformer architecture. As shown in Fig. 3, BATNet captures short-term fluctuations using wavelet transform and long-term dependencies with BiLSTM and Transformer, achieving high accuracy while maintaining interpretability and adaptability to real-time fluctuations.

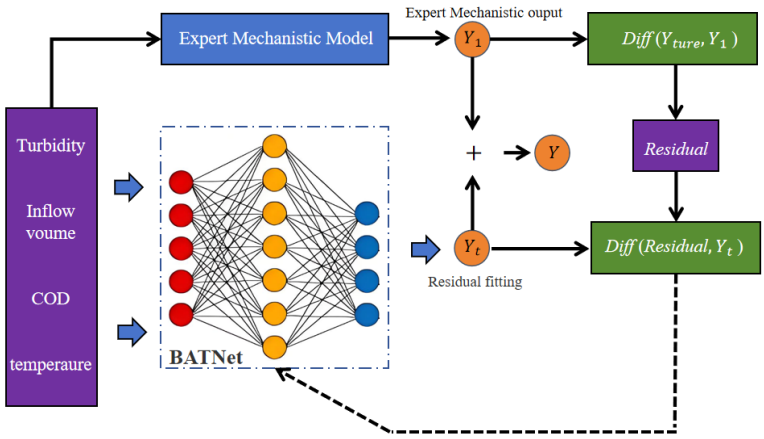


Fig. 3 – Prior Knowledge-Driven Expert Mechanistic Model with BATNet-Based Residual Fitting for Enhanced Dosing Prediction

Fig. 3 shows BATNet's workflow. Using key water-quality variables (e.g., turbidity, COD, and temperature), the expert model first provides an initial estimate Y_t^{cal}. The residual $r_t = Y_t^{\text{true}} - Y_t^{\text{cal}}$ is then modeled using BATNet, with wavelet transforms capturing short-term fluctuations and weighted attention guiding feature prioritization.

3.3 Bidirectional LSTM with Wavelet Transform for Enhanced Temporal Dependency Modeling

To capture complex temporal dynamics in alum dosing, BATNet integrates four components: (i) Wavelet Transform for feature extraction, (ii) Bidirectional LSTM for temporal dependency modeling, (iii) Transformer for dynamic feature prioritization, and (iv) Weighted Attention for expert knowledge integration. This design ensures robust and interpretable predictions under dynamic conditions.

Fig. 4 shows the overall architecture. Raw water quality parameters are processed through sequential and attention-based modules to generate accurate predictions.

To provide a systematic description of the prediction task, the proposed model is formulated the alum dosing prediction problem as follows. Let $\mathbf{x}_{t-L+1:t}$ denote a sequence of historical observations within a sliding window of length L, where each observation contains SCADA-monitored water quality variables (e.g., turbidity, COD, temperature, and flow rate). The goal is to predict the alum dosage at time t, denoted as y_t. Here, t denotes the current time index in the time series, and the sliding window $x_{t-L+1:t}$ contains the L most recent observations used for prediction.

In practice, an expert-based estimation y_t^{cal} is first obtained from empirical operational rules. The proposed BATNet model then learns the residual component $r_t = y_t^{\text{true}} - y_t^{\text{cal}}$, and produces the final dosing prediction by

$$\hat{y}_t = y_t^{\text{cal}} + \hat{r}_t. \tag{10}$$

This formulation enables the model to leverage both expert knowledge and data-driven learning for robust real-time dosing prediction.

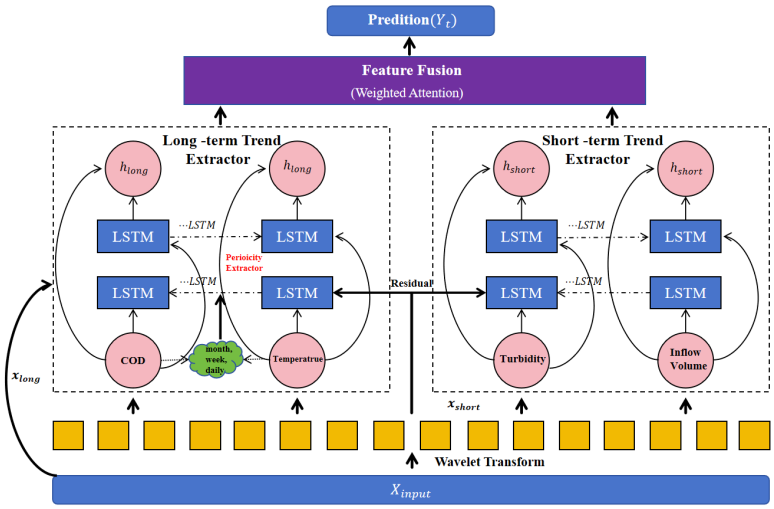


Fig. 4 – The structure of BATNet

To model both short-term fluctuations and long-term trends, the proposed model is applied wavelet transform to input signals, followed by a BiLSTM encoder to capture robust temporal dependencies. The BiLSTM enhances temporal representation, while online inference uses a sliding window of historical observations to prevent information leakage.

$$y_t = f\left(\tilde{\mathbf{x}}_t, \mathbf{h}_t^{\text{BiLSTM}}\right), \tag{11}$$

where $\tilde{\mathbf{X}}_t = [\tilde{A}_t, \tilde{B}_t, \tilde{C}_t]$ represents wavelet-transformed features at time t, and $\mathbf{h}_t^{\text{BiLSTM}}$ is the BiLSTM hidden representation.

By combining wavelet-based multiscale analysis with bidirectional temporal modeling, BATNet captures both abrupt disturbances and gradual variations, improving the stability and responsiveness of alum dosing predictions.

3.4 Transformer and Weighted Attention for Enhanced Dosing Predictions

Accurate alum dosing in water treatment systems requires not only understanding temporal dependencies but also prioritizing critical features and embedding domain-specific knowledge. The proposed model integrates a Transformer architecture with a Weighted Attention mechanism to achieve these goals. The Transformer focuses on dynamically assigning importance to features, while Weighted Attention incorporates prior expert knowledge to further refine dosing predictions.

For notational clarity, scalar quantities are denoted by italic symbols, vectors by bold lowercase symbols, and matrices by bold uppercase symbols throughout this paper. For example, y_t, r_t, and d_k denote scalars; $\mathbf{x}_t$, $\mathbf{w}_t$, $\mathbf{h}_t$, and $\boldsymbol{\alpha}_t$ denote vectors; and $\mathbf{X}$, $\mathbf{Q}$, $\mathbf{K}$, and $\mathbf{V}$ denote matrices.

3.4.1 Role of Transformer in Feature Prioritization

The Transformer architecture enhances the model's capability by dynamically assigning importance to different features and time steps based on their relevance to dosing outcomes. Its self-attention mechanism is particularly effective in identifying the critical interactions among water quality parameters, such as turbidity (A_t), COD (B_t), and temperature (C_t).

Let $\mathbf{H}_t \in \mathbb{R}^{L\times d_h}$ denote the hidden representation produced by the BiLSTM encoder. The corresponding query, key, and value matrices are then constructed as $\mathbf{Q}_t = \mathbf{H}_t\mathbf{W}_Q$, $\mathbf{K}_t = \mathbf{H}_t\mathbf{W}_K$, and $\mathbf{V}_t = \mathbf{H}_t\mathbf{W}_V$, where $\mathbf{Q}_t, \mathbf{K}_t \in \mathbb{R}^{L\times d_k}$ and $\mathbf{V}_t \in \mathbb{R}^{L\times d_v}$. The matrices $\mathbf{W}_Q \in \mathbb{R}^{d_h\times d_k}$, $\mathbf{W}_K \in \mathbb{R}^{d_h\times d_k}$, and $\mathbf{W}_V \in \mathbb{R}^{d_h\times d_v}$ denote trainable projection matrices.

The self-attention weight matrix and the Transformer output are computed as:

$$\begin{aligned} \mathbf{A}_t &= \text{softmax}\left(\frac{\mathbf{Q}_t \mathbf{K}_t^\top}{\sqrt{d_k}}\right), \\ \mathbf{Z}_t &= \mathbf{A}_t \mathbf{V}_t. \end{aligned} \tag{12}$$

Here, $\mathbf{A}_t \in \mathbb{R}^{L\times L}$ denotes the self-attention weight matrix, which measures the pairwise relevance among temporal features, and $\mathbf{Z}_t \in \mathbb{R}^{L\times d_v}$ denotes the output representation generated by the Transformer attention mechanism. The term d_k is the dimensionality of the key vectors and is used to scale the dot-product operation for numerical stability.

The output is a weighted combination of the input features, where the weights reflect the significance of each feature to the dosing prediction. By dynamically adjusting these weights, the Transformer enables the model to focus on the most relevant aspects of the input sequence, such as sudden spikes in turbidity or gradual changes in COD.

3.4.2 Weighted Attention for Domain Knowledge Integration

While the Transformer dynamically assigns feature importance, the Weighted Attention mechanism explicitly incorporates prior domain knowledge into the model. This ensures that critical water quality parameters, identified through expert insights, are given appropriate emphasis in the dosing predictions.

To incorporate prior operational knowledge, an expert-guided weight vector $\mathbf{w}_t \in \mathbb{R}^L$ is introduced. Let $\mathbf{a}_t$ denote the attention vector corresponding to time step t, derived from the self-attention matrix $\mathbf{A}_t$. The weighted attention vector is formulated as:

$$\tilde{\mathbf{a}}_t = \mathbf{w}_t \odot \mathbf{a}_t, \tag{13}$$

where $\mathbf{a}_t$ denotes the attention vector corresponding to time step t, and $\odot$ denotes element-wise multiplication.

3.4.3 Unified Architecture and Loss Design for Robust Dosing Prediction

To address the dynamic, multivariate nature of alum dosing, a unified architecture combining Bidirectional LSTM, Transformer, and Weighted Attention mechanisms is proposed. The model processes sequences of water quality variables, including turbidity (A_t), COD (B_t), temperature (C_t), and flow volume, to jointly capture temporal dependencies, feature importance, and domain knowledge.

The Bidirectional LSTM captures bidirectional temporal dependencies, producing a comprehensive representation:

$$\mathbf{h}_t^{\text{BiLSTM}} = \left[\mathbf{h}_t^{\text{forward}}; \mathbf{h}_t^{\text{backward}}\right], \tag{14}$$

The Transformer module dynamically assigns attention to critical features and time steps, generating attention weights α_t, which are further adjusted using expert-derived weights w_t via a Weighted Attention mechanism:

$$\boldsymbol{\alpha}_t^{\text{weighted}} = \mathbf{w}_t \odot \boldsymbol{\alpha}_t, \tag{15}$$

The final dosing prediction at time t integrates temporal representations and weighted attention:

$$\hat{y}_t = g\left(\mathbf{h}_t^{\text{BiLSTM}}, \tilde{\mathbf{a}}_t\right). \tag{16}$$

To enhance prediction accuracy while emphasizing the most relevant input variables, a composite loss function with dynamic sparsity adjustment is employed. The residual fitting is measured using mean squared error:

$$\mathscr{L}_{\text{MSE}} = \frac{1}{N}\sum_{i=1}^{N}(y_i^{\text{true}} - \hat{y}_i)^2 \tag{17}$$

while sparsity regularization is imposed on the feature attention weights to suppress irrelevant inputs:

$$\mathscr{L}_{\text{sparsity}} = \|\mathbf{A}\|_1 = \sum_{j=1}^{d} |a_j| \tag{18}$$

The total loss combines both terms:

$$\mathscr{L}_{\text{total}} = \mathscr{L}_{\text{MSE}} + \lambda_2 \mathscr{L}_{\text{sparsity}} \tag{19}$$

with a dynamically scheduled sparsity weight:

$$\lambda_2 = \min\left(\lambda_2^{\text{init}} + \alpha \cdot \text{epoch}, \lambda_2^{\max}\right) \tag{20}$$

This integrated design allows the model to initially focus on accurate residual prediction and progressively enforce sparsity, yielding robust, interpretable, and adaptive dosing predictions under complex, multivariate water quality conditions.

The model parameters are learned by minimizing the total loss function over the training data. Formally, the optimization problem of BATNet can be written as:

$$\theta^* = \arg\min_{\theta} \mathscr{L}_{\text{total}}(\theta), \tag{21}$$

where θ denotes all trainable parameters of the BATNet architecture, including the BiLSTM encoder, the Transformer attention module, the expert-guided feature gating module, and the prediction head. The parameters are optimized using backpropagation with the AdamW optimizer.

4. Model Evaluation

To validate the proposed BATNet framework introduced in Section 3, evaluations are conducted on two real-world water treatment datasets. The evaluation directly follows the problem formulation and model architecture described in Section 3, where the wavelet-enhanced BiLSTM–Transformer model is trained to predict alum dosing using historical water-quality observations. The source code and example datasets are available in an anonymous GitHub repository:`https://github.com/tina2020-liu/BATNet-dosing-prediction`. The repository includes the implementation of BATNet, training scripts, and evaluation code to facilitate reproducibility.

4.1 Training Settings

BATNet is trained end-to-end using the AdamW optimizer. The initial learning rate is set to 1×10^{-3}, the weight decay is 1×10^{-4}. In practice, the training process typically converges within about 40–50 epochs, where the validation RMSE stabilizes and no further significant improvement is observed. The batch size is 64 for training and 128 for validation. Gradient clipping with a maximum norm of 5.0 is adopted to stabilize optimization.

The sequence samples are split into training and validation subsets with a ratio of 8:2. The best model is selected according to the validation RMSE. The loss function consists of a mean squared error term and a sparsity regularization term imposed on the expert-guided gate. The sparsity coefficient is dynamically scheduled by:

$$\lambda_2 = \min\left(\lambda_2^{\text{init}} + \alpha \cdot \text{epoch}, \lambda_2^{\max}\right), \tag{22}$$

where $\lambda_2^{\text{init}} = 0.0$, $\lambda_2^{\max} = 10^{-3}$, and $\alpha = 10^{-5}$.

4.2 Comparison with Mainstream Algorithms

To benchmark the proposed BATNet, this study compares it with several representative mainstream models (XGBoost, BP, LSTM, and BiLSTM) on two real-world datasets, as summarized in Table 1. The baseline models differ in their modeling capabilities. BP is a feedforward neural network without explicit temporal modeling, while XGBoost captures nonlinear feature interactions but ignores sequential dependencies. LSTM models temporal dependencies in sequential data, and BiLSTM further incorporates bidirectional temporal context. In contrast, BATNet integrates wavelet-based multiscale decomposition, bidirectional temporal encoding, Transformer-based long-range interaction modeling, and expert-guided weighted attention, providing stronger adaptability to dynamic and nonlinear water-quality variations.

Table 1

Performance comparison of different dosing prediction models on datasets from Shanghai Xinghuo Water and Suzhou China-France Water.

Region	Model	Training set			Validation set			Test set			All sets		
		R^2	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE
Shanghai Xinghuo Water	XGBoost	0.962	0.110	0.154	0.945	0.125	0.171	0.934	0.132	0.182	0.952	0.118	0.164
	BP	0.915	0.162	0.229	0.886	0.181	0.252	0.874	0.191	0.265	0.901	0.173	0.242
	LSTM	0.954	0.121	0.168	0.936	0.137	0.186	0.926	0.144	0.197	0.944	0.130	0.179
	BiLSTM	0.966	0.109	0.151	0.951	0.122	0.169	0.942	0.128	0.178	0.958	0.116	0.162
	BATNet (Ours)	**0.988**	**0.063**	**0.089**	**0.982**	**0.071**	**0.099**	**0.976**	**0.078**	**0.108**	**0.984**	**0.068**	**0.095**
Suzhou China-France Water	XGBoost	0.954	0.124	0.172	0.936	0.140	0.191	0.924	0.148	0.203	0.944	0.133	0.183
	BP	0.898	0.182	0.255	0.866	0.203	0.282	0.852	0.214	0.296	0.882	0.194	0.269
	LSTM	0.946	0.137	0.191	0.926	0.155	0.211	0.914	0.163	0.224	0.936	0.147	0.203
	BiLSTM	0.959	0.123	0.172	0.942	0.138	0.189	0.931	0.146	0.201	0.950	0.132	0.181
	BATNet (Ours)	**0.984**	**0.072**	**0.102**	**0.977**	**0.081**	**0.114**	**0.971**	**0.088**	**0.124**	**0.979**	**0.078**	**0.110**

Table 1 compares BATNet with mainstream baselines (XGBoost, BP, LSTM, and BiLSTM) on two real-world datasets from Shanghai Xinghuo Water and Suzhou China-France Water. The models are evaluated using R^2, MAE, and RMSE on the same training, validation, and test splits to ensure a fair comparison. These metrics are widely used for regression evaluation: R^2 measures the goodness of fit between predictions and observations, MAE reflects the average prediction error, and RMSE is more sensitive to large deviations, which is particularly important for reliable dosing control.

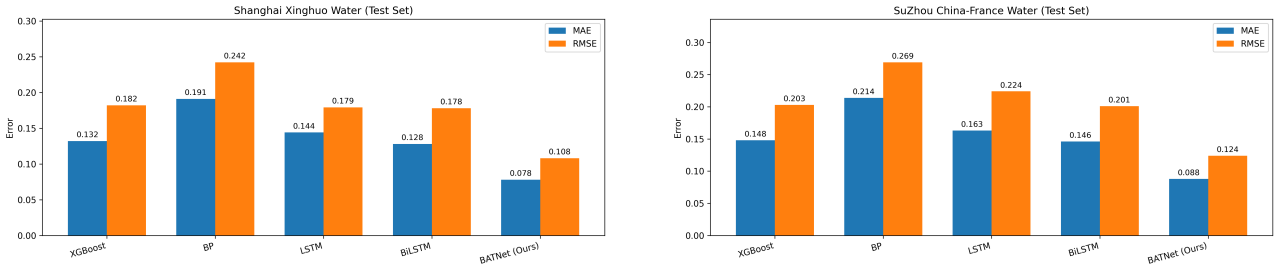
BATNet consistently achieves the best performance across both regions and all splits, indicating strong accuracy and generalization under dynamic, nonlinear, and multivariable water-quality conditions. On *Shanghai Xinghuo Water*, BATNet attains the best test results ($R^2 = 0.976$, MAE$= 0.078$, RMSE$= 0.108$). On *Suzhou China-France Water*, it remains superior ($R^2 = 0.971$, MAE$= 0.088$, RMSE$= 0.124$). In contrast, XGBoost/BP show weaker robustness to temporal dynamics, while LSTM/BiLSTM improve sequence modeling but remain limited in capturing long-range dependencies and adaptive feature re-weighting, leading to larger errors than BATNet.

Although BATNet achieves strong predictive performance, the potential risk of overfitting must be considered. To mitigate this issue, a strict training/validation/test split is adopted, and the validation set is used for model selection. Moreover, evaluations are conducted on two independent water treatment datasets (Shanghai Xinghuo Water and SuZhou China-France Water). The consistent performance improvements across both datasets suggest that the proposed framework generalizes well rather than memorizing patterns from a single dataset.

4.3 Visual Analysis of Test-Set Performance

To provide an intuitive comparison of predictive errors, Fig. 5 presents the MAE and RMSE results on the test sets of two real-world datasets, namely *Shanghai Xinghuo Water* and *Suzhou China-France Water*. BATNet consistently achieves the lowest MAE and RMSE on both datasets, demonstrating superior accuracy and robustness under dynamic and nonlinear water-quality variations. In contrast, conventional baselines, including XGBoost and BP, yield larger errors, while sequence models such as LSTM and BiLSTM improve predictive

performance but remain inferior to BATNet. These results further confirm the effectiveness of the proposed approach for reliable coagulant dosing prediction in practical scenarios.



(a) Shanghai Xinghuo Water (Test). (b) Suzhou China-France Water (Test).

Fig. 5 – MAE and RMSE comparison on the test sets for different models.

To further examine the model behavior at the sample level, Fig. 6 shows the comparison of normalized residual dosing amounts on the *Suzhou China-France Water* dataset. The predicted values follow the ground-truth trend closely across 300 samples, indicating that BATNet can effectively capture local fluctuations in dosing demand and maintain stable prediction accuracy under complex water-quality conditions. This observation is consistent with the quantitative results in Fig. 5 and further supports the practical applicability of the proposed approach.

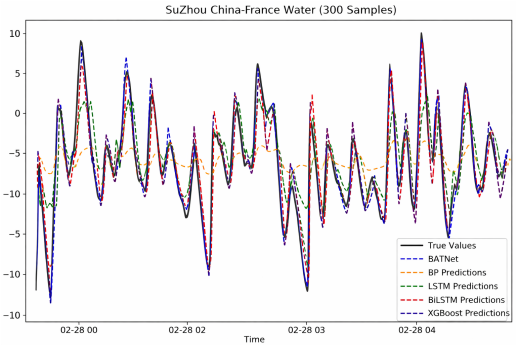


Fig. 6 – Comparison of normalized residual dosing amount on the *Suzhou China-France Water* dataset (300 samples).

4.4 Ablation Study

To quantify the contribution of each module in BATNet, this study conducts an ablation study, and the results are summarized in Table 2.

Table 2

Ablation study of BATNet on the test sets from Shanghai Xinghuo Water and Suzhou China-France Water.

Variant	**Shanghai Xinghuo Water(Test)**			**Suzhou China-France Water(Test)**		
	R^2	MAE	RMSE	R^2	MAE	RMSE
w/o WT (remove Wavelet Transform)	0.953	0.116	0.160	0.946	0.128	0.176
w/o BiLSTM (replace with LSTM)	0.959	0.109	0.151	0.952	0.121	0.168
w/o Transformer	0.947	0.124	0.171	0.938	0.138	0.189
w/o Weighted Attention	0.951	0.118	0.163	0.943	0.132	0.181
Full BATNet (Ours)	**0.976**	**0.078**	**0.108**	**0.971**	**0.088**	**0.124**

The full BATNet achieves the best results on both regional datasets, indicating that its performance gains arise from the complementary synergy of multiple modules rather than any single component. In particular, removing WT consistently degrades performance, confirming that wavelet-based multiscale decomposition

helps capture short-term/high-frequency fluctuations, mitigate non-stationarity, and provide more informative inputs for downstream temporal modeling. Replacing BiLSTM with a unidirectional LSTM increases errors, demonstrating the value of bidirectional temporal context in encoding richer sequence patterns and improving robustness to noisy operational variations. Removing the Transformer leads to a clear performance drop, highlighting the importance of self-attention in modeling long-range dependencies and complex feature interactions that are difficult to capture using recurrent structures alone. Finally, removing Weighted Attention also reduces accuracy, suggesting that expert-guided feature weighting improves feature prioritization, suppresses irrelevant disturbances, and stabilizes dosing predictions under varying and uncertain water-quality conditions.

5. Conclusion

This study proposes a prior knowledge-driven BiLSTM–Transformer model for accurate coagulant dosing prediction in dynamic water treatment systems. The framework integrates expert mechanistic priors with data-driven residual learning, preserving meaningful trends while adapting to nonlinearities and time-varying effects. Wavelet-based multiscale decomposition captures short-term fluctuations, BiLSTM models bidirectional temporal dependencies, and a Transformer captures long-range dependencies and cross-variable interactions. A weighted attention mechanism incorporates expert-guided feature prioritization to enhance interpretability and adaptability. Evaluations on two real-world datasets show that our approach outperforms conventional baselines in accuracy and stability, particularly under rapid changes. Future work will incorporate more domain knowledge and further validate the approach across diverse water treatment scenarios.

REFERENCES

[1] L. Li, S. Rong, R. Wang, and S. Yu, "Recent advances in artificial intelligence and machine learning for nonlinear relationship analysis and process control in drinking water treatment: A review," *Chemical Engineering Journal*, vol. 405, p. 126673, 2021.

[2] W. H. Pennock, M. L. Weber-Shirk, and L. W. Lion, "A hydrodynamic and surface coverage model capable of predicting settled effluent turbidity subsequent to hydraulic flocculation," *Environmental Engineering Science*, vol. 35, no. 12, pp. 1273–1285, 2018.

[3] B. Li, L. Liu, R. Ma, L. Guo, J. Jiang, K. Li, and X. Li, "Siamese based few-shot learning lightweight transformer model for coagulant and disinfectant dosage simultaneous regulation," *Chemical Engineering Journal*, vol. 499, p. 156025, 2024.

[4] S. Li, Y. Liu, Z. Wang, C. Dou, and W. Zhao, "Constructing a visual detection model for floc settling velocity using machine learning," *Journal of Environmental Management*, vol. 370, p. 122805, 2024.

[5] N. Wei, Z. Zhang, D. Liu, Y. Wu, J. Wang, and Q. Wang, "Coagulation behavior of polyaluminum chloride: Effects of pH and coagulant dosage," *Chinese Journal of Chemical Engineering*, vol. 23, no. 6, pp. 1041–1046, 2015.

[6] Q. Wei, J. Yang, F. Fu, and L. Xue, "Dynamic classification and attention mechanism-based bidirectional long short-term memory network for daily runoff prediction in Aksu River basin, Northwest China," *Journal of Environmental Management*, vol. 374, p. 124121, 2025.

[7] S. Li, Y. Liu, Z. Wang, C. Dou, W. Zhao, and H. Shu, "Characteristic analysis of floc size distribution and image texture evolution in chemical coagulation process," *Process Safety and Environmental Protection*, p. 107298, 2025.

[8] A. Bazer-Bachi, E. Puech-Coste, and J. L. Probst, "Mathematical modelling of optimal coagulant dose in water treatment plant," *Rev. Sci Edu*, vol. 3, no. 4, pp. 377–397, 1990.

[9] S. Haghiri, A. Daghighi, and S. Moharramzadeh, "Optimum coagulant forecasting by modeling jar test experiments using ANNs," *Drinking Water Engineering and Science*, vol. 11, no. 1, pp. 1–8, 2018.

[10] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[11] J. S. Jaganathan, S. R. S. Abdullah, S. N. A. Sanusi, N. N. Ramli, J. Alias, S. V. Subramaniam, N. M. Daud, F. A. Buslima, N. S. M. Said, J. Buhari, *et al.*, "Machine learning and explainable artificial intelligence in coagulation–flocculation: A contemporary review," *Journal of Environmental Chemical Engineering*, p. 119664, 2025.

[12] S. Peng, Y. Guo, J. Wang, Y. Wang, W. Zhang, X. Zhou, L. Jiang, and B. Lai, "The coagulation-precipitation turbidity prediction model for precision drug delivery system based on deep learning and machine vision," *Journal of Environmental Chemical Engineering*, vol. 12, no. 2, p. 112211, 2024.

[13] Y. Fang, Q. Sun, S. Wang, H. Wang, H. Xu, L. Huang, G. Li, Y. Li, and A. Wang, "Applicability of machine learning in predicting N_2O emission from wastewater treatment processes: a narrative review," *Journal of Environmental Management*, vol. 398, p. 128609, 2026.

[14] N. Jing, "Neural network-based pattern recognition in the framework of edge computing," *Romanian Journal of Information Science and Technology*, vol. 27, no. 1, pp. 106–119, 2024.

[15] L. Feng, Y. Zhang, X. Wei, M. Wang, Z. Niu, and C. Wang, "Interpretable prediction of coagulant dosage in drinking water treatment plant based on automated machine learning and SHAP method," *Journal of Water Process Engineering*, vol. 75, p. 107925, 2025.

[16] M. Achite, S. Samadianfard, N. Elshaboury, and M. Sharafi, "Modeling and optimization of coagulant dosage in water treatment plants using hybridized random forest model with genetic algorithm optimization," *Environment, Development and Sustainability*, vol. 25, no. 10, pp. 11189–11207, 2023.

[17] S. Faramarzmanesh, S. Mozaffari, S. E. H. Garmdareh, M. Mashal, and E. Ghane, "Development of PSO-SVR and ANN models to evaluate the effect of cross-sectional shape on denitrifying bioreactors," *Water Supply*, vol. 25, no. 11, pp. 1482–1497, 2025.

[18] C. E. Richards, A. Tzachor, S. Avin, and R. Fenner, "Rewards, risks and responsible deployment of artificial intelligence in water systems," *Nature Water*, vol. 1, no. 5, pp. 422–432, 2023.

[19] S. Wang, H. Peng, and S. Liang, "Prediction of estuarine water quality using interpretable machine learning approach," *Journal of Hydrology*, vol. 605, p. 127320, 2022.

[20] Y. Tang, Y. Wang, C. Liu, X. Yuan, K. Wang, and C. Yang, "Semi-supervised LSTM with historical feature fusion attention for temporal sequence dynamic modeling in industrial processes," *Engineering Applications of Artificial Intelligence*, vol. 117, p. 105547, 2023.

[21] S. Lin, J. Kim, C. Hua, M.-H. Park, and S. Kang, "Coagulant dosage determination using deep learning-based graph attention multivariate time series forecasting model," *Water Research*, vol. 232, p. 119665, 2023.

[22] M. Sharafi, V. Rezaverdinejad, J. Behmanesh, and S. Samadianfard, "Development of long short-term memory along with differential optimization and neural networks for coagulant dosage prediction in water treatment plant," *Journal of Water Process Engineering*, vol. 65, p. 105784, 2024.

[23] W. Zhao, Y. Liu, R. Shuang, W. Lu, S. Li, C. Zhao, C. Dou, and H. Shu, "Regulation of coagulant dosage in water treatment based on explainable integrated time-series deep learning models," *Process Safety and Environmental Protection*, p. 107613, 2025.

[24] A. O. Bankole, R. Moruzzi, R. G. Negri, J. Bridgeman, and S. Sharifi, "Image-based machine learning applications for flocculation modelling in water treatment: Prospects towards automation," *Journal of Hazardous Materials Advances*, p. 100870, 2025.

[25] J. Li and L. Zheng, "AQUATICGAN: An approach for enhancing underwater images," *Proceedings of the Romanian Academy, Series A: Mathematics, Physics, Technical Sciences, Information Science*, vol. 26, no. 4, 2025.

[26] J. T. Hancock, T. M. Khoshgoftaar, and Q. Liang, "A problem-agnostic approach to feature selection and analysis using SHAP," *Journal of Big Data*, vol. 12, no. 1, p. 12, 2025.

Received January 27, 2026