Water Research X 26 (2025) 100291

Contents lists available at ScienceDirect

Water Research X

journal homepage: www.sciencedirect.com/journal/water-research-x

A probabilistic deep learning approach to enhance the prediction of wastewater treatment plant effluent quality under shocking load events

Hailong Yin [a,b,*], Yongqi Chen [a,b], Jingshu Zhou [a,b], Yifan Xie [c], Qing Wei [a,b], Zuxin Xu [a,b,*]

[a] Shanghai Institute of Pollution Control and Ecological Security, Shanghai, 200092, China
[b] Key Laboratory of Urban Water Supply, Water Saving and Water Environment Governance in the Yangtze River Delta of Ministry of Water Resources, State Key Laboratory of Pollution Control and Resource Reuse, Tongji University, Shanghai, 200092, China
[c] School of Environment, Tsinghua University, Beijing, 100084, China

ARTICLE INFO

Keywords:
Wastewater treatment plant
Machine learning model
Deep learning
Effluent water quality
Shocking load

ABSTRACT

Sudden shocking load events featuring significant increases in inflow quantities or concentrations of wastewater treatment plants (WWTPs), are a major threat to the attainment of treated effluents to discharge quality standards. To aid in real-time decision-making for stable WWTP operations, this study developed a probabilistic deep learning model that comprises encoder-decoder long short-term memory (LSTM) networks with added capacity of producing probability predictions, to enhance the robustness of real-time WWTP effluent quality prediction under such events. The developed probabilistic encoder-decoder LSTM (P-ED-LSTM) model was tested in an actual WWTP, where bihourly effluent quality prediction of total nitrogen was performed and compared with classical deep learning models, including LSTM, gated recurrent unit (GRU) and Transformer. It was found that under shocking load events, the P-ED-LSTM could achieve a 49.7% improvement in prediction accuracy for bihourly real-time predictions of effluent concentration compared to the LSTM, GRU, and Transformer. A higher quantile of the probability data from the P-ED-LSTM model output, indicated a prediction value more approximate to real effluent quality. The P-ED-LSTM model also exhibited higher predictive power for the next multiple time steps with shocking load scenarios. It captured approximately 90% of the actual over-limit discharges up to 6 hours ahead, significantly outperforming other deep learning models. Therefore, the P-ED-LSTM model, with its robust adaptability to significant fluctuations, has the potential for broader applications across WWTPs with different processes, as well as providing strategies for wastewater system regulation under emergency conditions.

1. Introduction

Worldwide, the operations of WWTPs are challenged by abnormal events arising from sudden significant increases in WWTP inflow quantities and concentration, which may lead to over-limit concentration of WWTP effluent discharge (Karamnia et al., 2024). Reliable early-warning of the WWTP effluent concentration in response to high variability in WWTP inflow rate and concentration, is the basis for regulatory changes in sewer network and WWTP operations, for example, temporary operating down of the WWTP or reduction of WWTP inflow rate. To address this challenge, a robust and accurate prediction of WWTP effluent water quality is of great importance to aid in real-time decision-making for stable WWTP operations.

In the past years, with the development of artificial intelligence, machine learning has emerged as a promising tool for addressing complex non-linear problems in real-world applications (Huang et al., 2021; Zhong et al., 2021; Zhu et al., 2022). Traditional machine learning algorithms including artificial neural networks (ANNs) (Cimen Mesutoglu and Gok, 2024; Maurya et al., 2021; Mjalli et al., 2007; Xie et al., 2022), support vector regression (Guo et al., 2015; Nourani et al., 2021), extreme gradient boosting (Ching et al., 2022; Li et al., 2024), random forest (Manav-Demir et al., 2024; Sharafati et al., 2020), and perturbation theory machine learning (Badran et al., 2019; Ocampo-Perez et al., 2016), have been used for monitoring the operation of the wastewater treatment facilities. Application scenarios concerned the prediction of pollutant removal and management performance of a variety of physical, chemical, and biological treatment technologies, as well as effluent quality prediction in actual WWTPs for their real-time early-warning

* Corresponding authors.
E-mail addresses: yinhailong@tongji.edu.cn (H. Yin), xzx@tongji.edu.cn (Z. Xu).

https://doi.org/10.1016/j.wroa.2024.100291
Received 13 September 2024; Received in revised form 1 December 2024; Accepted 2 December 2024
Available online 3 December 2024

and supervision.

Besides the above models, with advancements in algorithms, researchers continue to propose machine learning approaches with stronger nonlinear mapping ability. Especially, the emergence of deep learning approaches is receiving increased attention (Huang et al., 2023). Some deep learning models such as convolutional neural networks (CNN) (Hu et al., 2024), long short-term memory (LSTM) (Li et al., 2023a; Liu et al., 2021; Zhang et al., 2023), gated recurrent unit (GRU) (Cheng et al., 2020; Jiang et al., 2021; Yan et al., 2023), temporal convolutional networks (TCN) (Li et al., 2023b), and Transformer (Chang et al., 2024), have been employed in WWTP operation management. Apart from the stand-alone deep learning models, hybrid deep learning models such as CNN-LSTM and TCN-LSTM, have also been employed in WWTP influent mass loading or effluent water quality prediction, aiding in real-time management strategies (Wang et al., 2019; Xie et al., 2024). Generally, deep learning models with enhanced neural network structure can bring a stronger ability to extract and memorize data characteristics, thus offering higher prediction accuracy than the traditional machine learning models, especially for the real-time predictions in actual WWTPs with complex operation scenarios.

However, these machine learning models for sewage treatment processes are primarily built upon a modeling framework to fully memorize and interpolate pre-set or prior knowledge. Under unpredicted influent shocking load events that may be beyond this prior knowledge, the prediction accuracy of these models would be undermined due to their deficiency in extrapolating the prior knowledge to future predictions with large uncertainty (Alvi et al., 2023). To address this challenge, recent methods for probabilistic modeling have emerged, offering probability distributions instead of deterministic values to assess uncertainties (Suresh et al., 2022, Zhang et al., 2022, Zheng and Zhang, 2024). However, these studies did not concern the use of probabilistic models in unpredicted extreme cases with large uncertainty (i.e., shocking load events in this case). A robust probability modeling is still anticipated to improve the prediction accuracy under shocking load events (Jaffari et al., 2024; Kim et al., 2019).

Currently, in WWTP operations, total nitrogen (TN) is the most challenging indicator for the attainment of prescribed effluent quality standards (Sadri Moghaddam and Mesghali, 2023; Wang et al., 2017). Excessive discharge of TN into surface waters may cause severe harm to the aquatic environment and human health (e.g., methemoglobinemia) (Suresh et al., 2023). Due to the complicated interactions of WWTP influent and operational parameters, effluent TN concentration would fluctuate considerably under shocking load scenarios, making its accurate predictions more difficult.

Motivated by above challenges, this study developed a novel probabilistic encoder-decoder LSTM (P-ED-LSTM) modeling framework to enhance the WWTP effluent quality prediction. We explored the robustness of the proposed model and compared it with the classical deep learning models in an actual WWTP, where influent shocking load events once occurred.

2. Results and discussions

2.1. Comparison of the performance between the P-ED-LSTM model and other deep learning models

The performance of the developed P-ED-LSTM model (see Section 4.1) was compared with those of the classical deep learning models including the traditional LSTM, GRU, and Transformer models (Supporting Information S2-S4). For the whole dataset (i.e., the data from July 1, 2022 to June 30, 2023), comparisons of the predicted effluent TN accuracies among the four deep learning models are demonstrated in Fig. 1 and Table 1～Table 2, which are based on 2 h ahead (i.e., one time-step ahead) prediction. Additional information for the effluent quality prediction was provided in the Supporting Information S5, highlighting that the following discussions are universal for effluent quality prediction of different parameters.

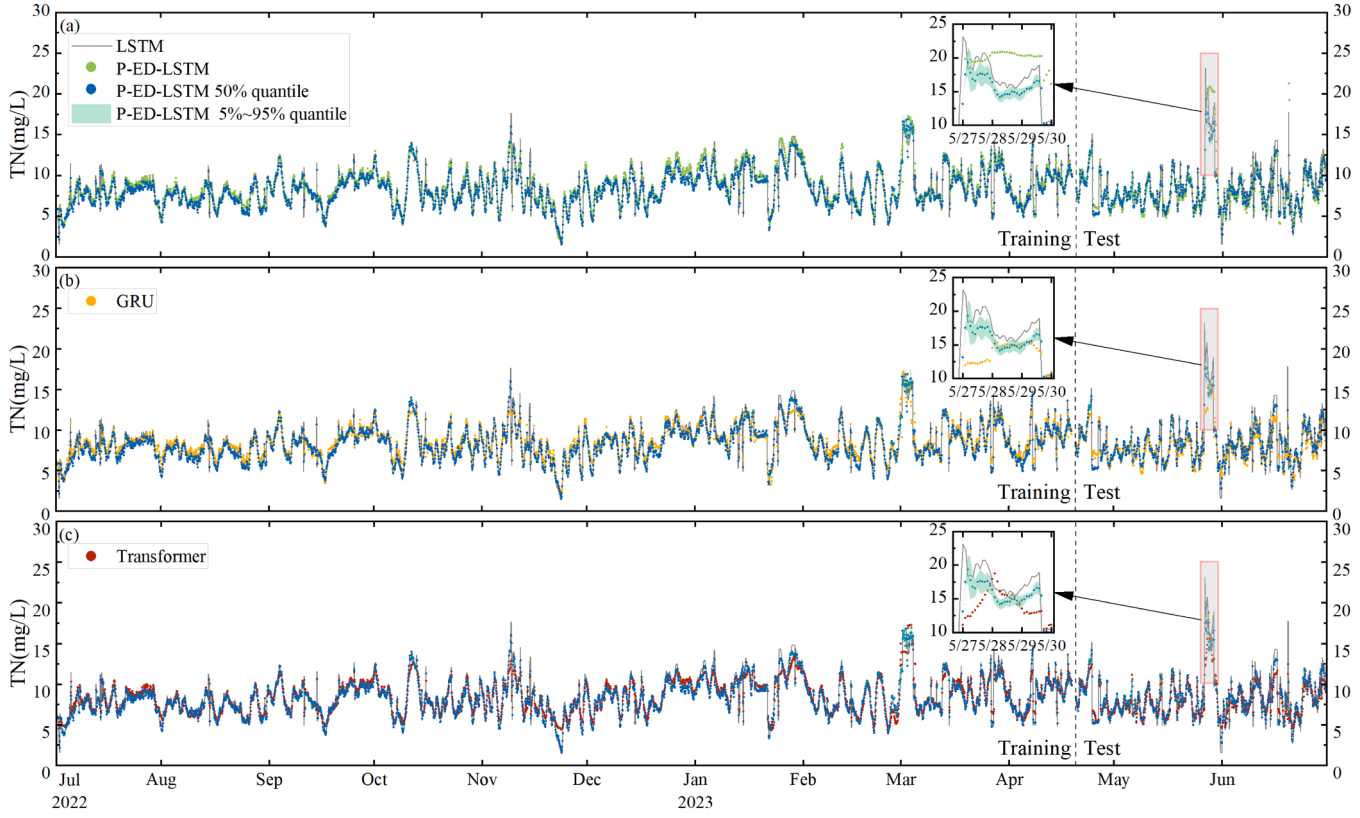

Fig. 1. Comparison of time-series predictions between the P-ED-LSTM model and the other deep learning models of (a) LSTM, (b) GRU, and (c) Transformer model.

Specifically, Fig. 1 offers comparisons of time-series predictions between the P-ED-LSTM model and the other models, for both the training and test dataset. Tables 1 and 2 quantify the modeling performance using four metric indicators of mean absolute percentage error (MAPE), mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination (R^2), as explained in Eqs. (6)-(9). For the P-ED-LSTM model, prediction accuracy under 50% quantile estimate is provided in this Table. Theoretically, the lower MAPE, MAE, and RMSE, and the higher R^2 throughout the data groups, are indications of the modeling's robustness.

Fig. 1 and Table 1 show that for the training dataset, the performance of the developed P-ED-LSTM model is close to that of LSTM model. However, if assessing by the test dataset, the P-ED-LSTM model outperforms the LSTM model. Clearly, this demonstrates that the P-ED-LSTM model can circumvent the overfitting problem of LSTM model. Fig. 1 and Table 1 also show the P-ED-LSTM model outperform the GRU and Transformer model for both training and test dataset. Table 2 further compares the performances of different models for the test dataset under influent shocking loads and other WWTP operation scenarios, respectively. It reveals that the improvement in prediction accuracy with the P-ED-LSTM model is due to better agreement for total observations instead of partial observations in test dataset. Recent studies (Liu et al., 2023; Lv et al., 2024; Yang et al., 2022) reported the effluent TN prediction in actual WWTPs; however, the WWTP effluent TN concentrations were not beyond the prescribed discharge standard, and no exceptional cases such as over-limit discharges were included in modeling prediction. By contrast, our proposed model has a robust prediction for a wide range of effluent quality including the potential over-limit discharge due to influent shocking load.

Advantage of the P-ED-LSTM model over the other three deep learning models, can be firstly explained by designed encoder-decoder learning structure of the developed model (see Fig. 5). By introducing this novel structure, the P-ED-LSTM model excels at capturing complex time-series data relationships. The encoder processes the input data sequence, emitting contextual information that represents global historical trends and dependencies, while the decoder utilizes this contextual information to generate forecasts (Kao et al., 2020; Wang et al., 2021). In this way, the P-ED-LSTM model showed higher accuracy than the LSTM and GRU model. It was noted that similar encoder-decoder structure was also adopted by the Transformer model. However, its self-attention mechanism inherently fails to preserve the temporal order of time series data (i.e., temporal information loss) (Zeng et al., 2023), which could explain the lower performance of the Transformer model compared to the P-ED-LSTM model. Further, by adding the capability of producing probability predictions, the P-ED-LSTM model demonstrated its superior performance for real-time effluent quality prediction under shocking load events, as discussed below.

2.2. Performance enhancement by the P-ED-LSTM model under shocking load events

2.2.1. Performance enhancement assessment with the metric indicators

Fig. S2 signifies continuous four-day influent shocking loads during May 27 to May 30, 2023, when the influent concentrations varied within 1.9~2.3 times of the averaged data under other WWTP operation cases. Correspondingly, the WWTP experienced significant increase in effluent TN concentration that exceeded the discharge limit (i.e., 15 mg/L according to Grade 1-A of China's GB18918–2002), as seen in the zoomed-up subplots of Fig. 1. Matching the on-line data collection frequency of 2 h (see Section 2.1), these shocking loads corresponded to 48 cases (i.e., 96 hours / 2 hours=48 cases), covering 5.5% of the entire test dataset.

Under shocking load events, compared to the classical models that can only offer definite point prediction, the developed P-ED-LSTM enables the output of prediction range under different confidence intervals for each time step, which extends the search of predicted data approximate to the real effluent data. In this way, the P-ED-LSTM model can offer higher generalization ability than the other deep learning models.

Concretely, for the LSTM model, it produced significantly higher effluent TN concentration than the observed data under shocking load scenarios (Fig. 1a). Moreover, its prediction output seemed stable, failing to capture the real effluent situations with large variations. For the GRU and Transformer models, they provided lower effluent TN predictions than the observed data under shocking load events (Fig. 1b and Fig. 1c), which may fail to identify the potential over-limit discharges due to under-estimated predictions. In contrast, the PED-LSTM model produced time-varying predictions close to the real effluent situations, highlighting its performance in capturing highly transient future trends through full exploration of contextual information with added likelihood prediction.

Table 2 presents quantitative evidence of improved WWTP effluent prediction accuracy for shocking load events with the P-ED-LSTM model. It shows that a higher quantile of the probability data generated from the P-ED-LSTM model, indicates a predicted value more approximate to the real effluent quality. R^2 of the P-ED-LSTM model steadily increased from 0.598 to 0.803 when the probability data quantile increased from 5% to 95%. With the 95% quantile estimate, the P-ED-LSTM model could achieve 49.7%~57.6% higher accuracy than the LSTM, GRU, and Transformer models, when measured by the MAE, MAPE, and RMSE. The R^2 significantly increased by 2.7 times after the introduction of the P-ED-LSTM model.

2.2.2. Performance enhancement assessment with the Taylor diagram

The Taylor diagram was introduced to measure the overall performance of the above deep learning models under shocking load events, as shown in . This figure shows four model points associated with the P-ED-LSTM, LSTM, GRU, and Transformer, which were located by the calculated three metrics of correlation coefficient, standard deviation, and centered root mean square error (CRMSE) (see Eqs. (11)-(14)). There is also a generalized measurement point in this figure, with correlation coefficient of 1.0, standard deviation of 4.81, and CRMSE of 0, respectively.

Theoretically, the closer the model point is to the measurement point, the better the model's overall performance is. Clearly, the P-ED-LSTM model corresponded to a model point with the closest distance to the measurement point. Specifically, the distances between the models and measurement point were 2.016, 3.473, 3.537, and 3.399 for the P-ED-LSTM, LSTM, GRU, and Transformer, respectively. Therefore, the overall performance for the effluent quality prediction under shocking load events was improved by 40.7%~43.0% after the introduction of the P-ED-LSTM model.

From a wider perspective, Fig. 2 also reveals the similarities among the models. Specifically, the narrow angle between the lines in Fig. 2

Table 1
Metrics of prediction accuracies of the four deep learning models for the whole training and test dataset.

Deep learning models	Training dataset				Test dataset			
	MAE	MAPE	RMSE	R^2	MAE	MAPE	RMSE	R^2
LSTM	0.457	0.057	0.836	0.886	0.954	0.120	1.669	0.679
GRU	0.615	0.080	0.917	0.863	1.050	0.133	1.667	0.679
Transformer	0.670	0.086	1.091	0.806	1.009	0.126	1.597	0.706
P-ED-LSTM (50% quantile)	0.482	0.058	0.811	0.893	0.729	0.091	1.224	0.827

Table 2
Metrics of prediction accuracies of the four deep learning models for the test dataset under influent shocking load events and other scenarios.

Deep learning models	Dataset under influent shocking load events				Dataset excluding influent shocking load events			
	MAE	MAPE	RMSE	R^2	MAE	MAPE	RMSE	R^2
LSTM	3.178	0.233	4.192	0.223	0.823	0.113	1.384	0.791
GRU	3.050	0.177	4.406	0.142	0.932	0.130	1.341	0.804
Transformer	2.873	0.171	4.000	0.293	0.899	0.123	1.326	0.808
P-ED-LSTM (5% quantile)	2.379	0.145	3.017	0.598	0.669	0.089	1.115	0.865
P-ED-LSTM (25% quantile)	1.991	0.124	2.652	0.689	0.668	0.089	1.113	0.865
P-ED-LSTM (50% quantile)	1.744	0.110	2.421	0.741	0.669	0.090	1.114	0.865
P-ED-LSTM (75% quantile)	1.515	0.098	2.240	0.778	0.671	0.090	1.116	0.864
P-ED-LSTM (95% quantile)	1.283	0.085	2.110	0.803	0.676	0.091	1.124	0.862

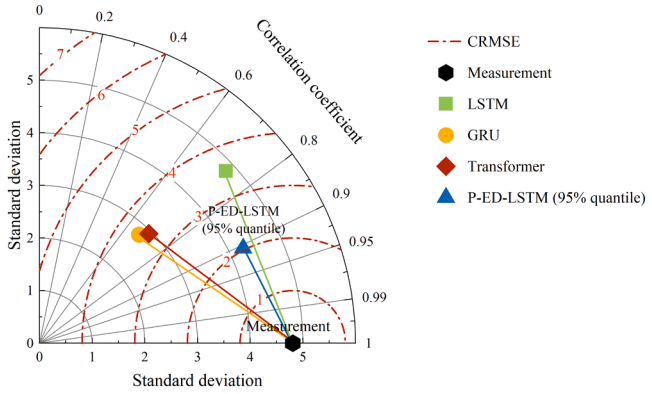

Fig. 2. The Taylor diagram for assessing the overall performance of the P-ED-LSTM and other deep learning models under shocking load events. Correlation coefficient, standard deviation, and CRMSE are marked in this figure.

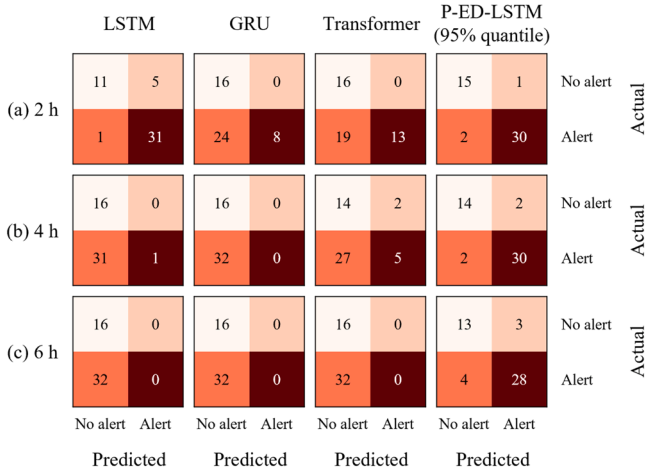

Fig. 3. Counts of the alerts forecasted by the P-ED-LSTM and other deep learning models under shocking load events, with early-warning time of (a) 2 h, (b) 4 h, and (c) 6 h.

signifies a positive and close correlation. It shows that the GRU model exhibits a strong similarity with the Transformer model, which was also reported by the recent studies from Hou et al. (2024) and Xu et al. (2023). As aforementioned, both the GRU and Transformer model feature underestimation of the actual effluent data under shocking load scenarios (Fig. 1b and Fig. 1c). For the P-ED-LSTM model, it correlates well with the LSTM model, which can be explained by the fact that the P-ED-LSTM model is an enhanced LSTM model comprising encoder-decoder LSTM networks with added capacity of producing probability predictions, as depicted in Fig. 5. After improving the LSTM framework, the P-ED-LSTM model offers a robust prediction that matched well with the high-transient WWTP effluent TN concentration.

2.3. Extending the P-ED-LSTM model to multi-time-step ahead prediction

To further test the robustness and serviceability of the proposed P-ED-LSTM model, it was also applied to multi-time-step ahead prediction to provide early-warning decision-making under shocking load events. For comparison purposes, the early-warning time was set to three cases of ahead 2 h, 4 h, and 6 h, which matched ahead one, two, and three prediction time steps. Two levels including no alert and alert were set, corresponding to effluent TN concentration less than or equal to 15.0 mg/L and greater than 15.0 mg/L, respectively. Alert indicated over-limit discharge of Grade 1-A standard (China's GB18918–2002).

Fig. 3 presents the comparison between the actual and predicted alerts under shocking load scenarios, where the predicted alerts were given by the P-ED-LSTM (95% quantile output), LSTM, GRU, and Transformer, respectively. Totally 48 cases were included in this figure as explained in Section 2.2.1.

With the P-ED-LSTM, the early-warning accuracy (see Eq. (10)) under ahead time of 2 h, 4 h, and 6 h were 93.8%, 91.7%, and 85.4%, respectively. When employing the models of LSTM, GRU, and Transformer, the early-warning accuracy declined obviously. Under ahead prediction time of 2 h, the early-warning accuracy by the LSTM, GRU, and Transformer were 87.5%, 50%, and 60.4%, respectively. The early-warning accuracies with the three models sharply declined to 33.3% ~39.6% under ahead prediction time of 4 h~6 h. Specially, when early-warning time was extended to 6 h, the LSTM, GRU, and Transformer failed to predict the actual over-limit discharges. By contrast, the developed P-ED-LSTM model still successfully captured 87.5% of the actual over-limit discharges (i.e., predicted 28 cases versus actual 32 cases in Fig. 3c), further demonstrating its superior performance in capturing long-term temporal dependencies with large variations.

2.4. Further understanding the P-ED-LSTM model through SHAP analysis

To further explore the intrinsic mechanism of the proposed model, SHAP analysis was conducted to understand the contributions of different input features to the prediction results (Fig. 4). As stated in Section 4.5.3, a larger SHAP value represents a greater contribution of the modeling input parameters to the effluent quality. Fig. 4 provides the calculated importance of each input variable from high to low-ranking measuring by the SHAP values, which shows effluent and influent parameters are more influential than the process monitoring parameters.

Furthermore, a decreasing experiment was conducted by removing influent parameters and process controlling parameters, respectively. The results of the decreasing experiments are shown in Supporting Information S6. It shows that the WWTP influent parameters are more influential than process monitoring parameters. This implies that alleviating the influent shocking load (e.g., coordinated regulation between sewer network and WWTP) instead of optimizing WWTP treatment processes, is the basic measure to assure the stable WWTP operation and attainment of effluent quality to the discharge standard under extreme

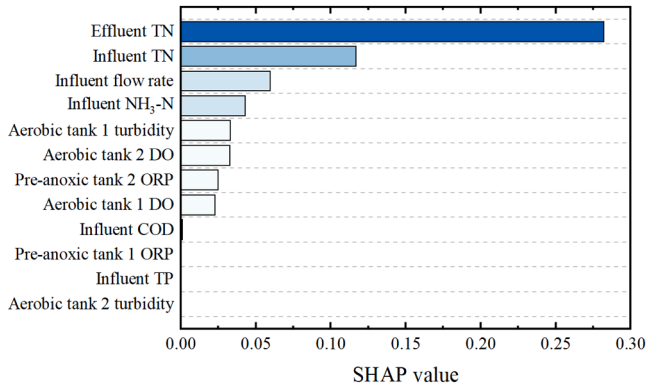

Fig. 4. Ranking of the importance of model input parameters through SHAP analysis.

situations.

3. Conclusions

The designed P-ED-LSTM model with its encoder-decoder structure can capture the intricate relationship of the time-series data. Additionally, the P-ED-LSTM model's capability of producing probability predictions significantly enhances its capability to extrapolate prior knowledge to extreme scenarios, particularly under influent shocking load events.

Both the metric indicators and the early-warning accuracy rates for the over-limit discharges demonstrate that the P-ED-LSTM is superior to the LSTM, GRU, and Transformer. The P-ED-LSTM model achieved a 49.7% improvement in prediction accuracy for bihourly predictions of WWTP effluent TN concentration compared to the other models. Also, the P-ED-LSTM model captured approximately 90% of the actual over-limit discharges up to 6 hours ahead, significantly outperforming the other deep learning models.

However, as limited data for shocking load events were available in this study, the P-ED-LSTM model can be further tested with more shocking load cases in different WWTPs. Additionally, the current P-ED-LSTM model does not incorporate an optimal control module for minimizing energy consumption while ensuring compliance with effluent standards. Therefore, in the future, the PED-LSTM model can be integrated with optimal control algorithms to further optimize WWTP regulation under emergency conditions.

4. Materials and methods

4.1. Development of the P-ED-LSTM model

The developed novel deep learning model comprising encoder-decoder LSTM networks with added capacity of producing probability predictions (i.e., P-ED-LSTM model) was proposed for WWTP effluent TN prediction under shocking load events. In detail, the structure of the P-ED-LSTM model is illustrated in Fig. 5, which consists of input layer, LSTM layer, dense layer, probability model layer, and output layer in turn. The model was developed with the open-source deep-learning framework of PyTorch in Python 3.7.

In Fig. 5a, the data input into the model for WWTP effluent quality prediction comprises historical time-series information of WWTP influent parameters and process monitoring parameters. In this case, the WWTP influent parameters include lifting pump inflow, influent water quality data for COD, ammonia (NH_3-N), total phosphorus (TP), and TN; the WWTP process monitoring parameters include two oxidation–reduction potential (ORP) meters in pre-anoxic tanks, two dissolved oxygen (DO) meters, and two turbidity meters in aerobic tanks.

Currently, the on-line instruments at the WWTP outlet are mandatory for real-time monitoring of effluent water quality. Considering the historical on-line effluent quality data are easily available, the model input also incorporates historical effluent TN data to capture temporal dependencies among the effluent data sets and then enhances the inference of future effluent quality trends.

Specifically, the data input is arranged by a method of sliding window, with which the model output at the time t corresponds to the sequential data input from the previous time periods of $t-l\times k$ to $t-k$. In this case, k is the prediction time interval which is set to 2 hours that matches the online data collection frequency for the studied WWTP. l is set to a value of 13 so as to match the hydraulic retention time of 26 hours for the studied WWTP (i.e., $l\times k = 26$ hours). Hence, the model output at prediction time t is associated with the input of previous time periods from $t-26$ to $t-2$, which corresponds to the input data sequence from $i-13$ to $i-1$ denoted by matrix X_i in Fig. 5a.

The WWTP effluent TN water quality is predicted following the steps below.

(1) The acquired WWTP data are divided into a training dataset and a test dataset. With the data input described above, an encoder-decoder architecture in combination with the LSTM network (see Supporting Information S2 for more details) is designed to establish a learning model that can capture the intricate relationship of the time-series data. The architecture comprises an encoder, a state vector, and a decoder (see Fig. 5b).

Specifically, at prediction time t, the encoder reads the input data sequence from $i-13$ to $i-2$ (corresponding to the historical time periods from $t-26$ to $t-4$) and converts it into a fixed-length vector with a hidden cell. The decoder then uses this fixed length vector along with the data sequence of $i-1$ (corresponding to the previous time at $t-2$), to produce the hidden state h_i.

(2) The dense layer further processes the hidden state h_i to calculate the mean value μ and standard deviation value σ following the mathematical expressions below:

$$\mu = w_\mu h_i + b_\mu \tag{1}$$

$$\sigma = \log(1+\exp(w_\sigma h_i + b_\sigma)) \tag{2}$$

where w_* and b_* are the weight and bias of the dense layer, respectively.

(3) With the determined μ and σ, in the mode training stage, a Gaussian likelihood function is produced in the probability model layer based on the following expression:

$$p\left(y_{\text{meas},i}|\mu,\sigma\right) = \left(2\pi\sigma^2\right)^{-\frac{1}{2}}\exp\left(-\frac{\left(y_{\text{meas},i}-\mu\right)^2}{2\sigma^2}\right) \tag{3}$$

where $y_{\text{meas},i}$ is measured effluent TN of the data sequence i (corresponding to the time at t) of the training datasets.

Then, the modeling parameters are determined by minimizing the loss function L of log-likelihood form that is described below:

$$L = -\sum_{i=1}^{n}\log p\left(y_{\text{meas},i}|\mu,\sigma\right) \tag{4}$$

where n is the number of predicted data series. In this way, the model is optimally tuned and the model training is finished.

(4) After the model training, the test dataset is used as input to further verify the developed model. In this stage, a probabilistic forecast $y_{\text{pred},i}$ is produced following the Gaussian distribution in the probability model layer, and accordingly consistent quantile estimates for all subranges in the predicted horizon can be provided in the output layer (see Fig. 5b).

In more detail, the hyperparameters for model running include one LSTM layer with 320 neurons and a dropout rate of 0.5. The model training is performed based on the Adam optimizer, and the learning rate is set as 0.005. The model runs for 100 epochs with a batch size of

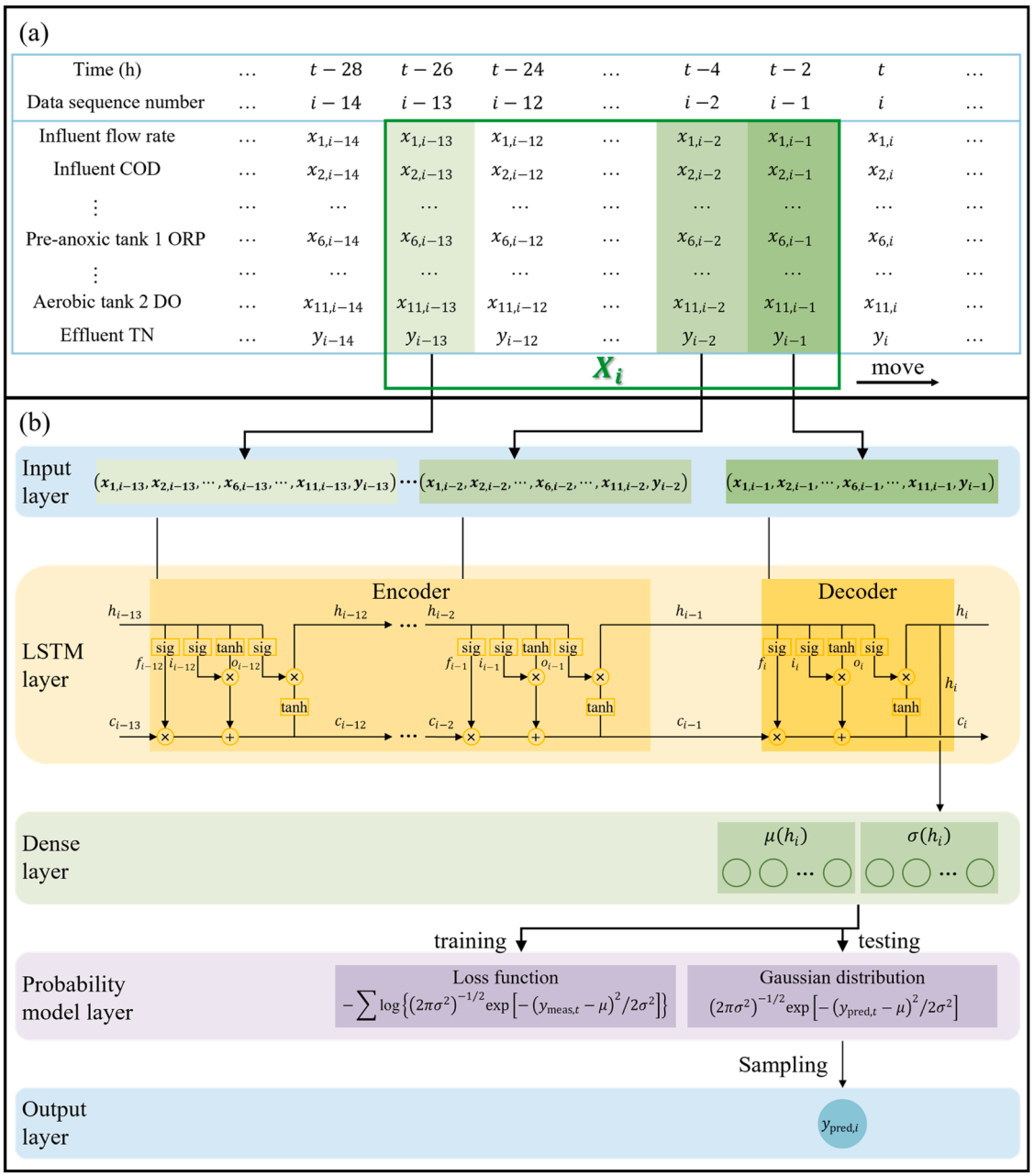

Fig. 5. Design of the P-ED-LSTM model for WWTP effluent prediction. (a) The sampling strategy using the sliding window method. (b) The framework of the probabilistic model. h_i and c_i are the hidden state and internal state; sig and tanh represent the sigmoid and hyperbolic tangent activation functions; f_i, i_i, and o_i are the forget, input, and output gates, respectively.

32, and a patience setting of 20 for early stopping to mitigate overfitting.

4.2. The deep learning models for comparison: LSTM, GRU, and transformer

LSTM is a modified version of recurrent neural network (RNN) architecture for time-series predictions (Van Houdt et al., 2020). It consists of multiple memory cells with the hidden state and the cell state in each of the cell. The hidden state is responsible for short-term memory, and the cell state has the capacity for long-term memory. In addition, each memory cell consists of three internal gates (i.e., forget, input, and output gates), which process the information by determining which information needs to be removed or saved in the cell for learning long-term temporal dependencies. A more detailed introduction to LSTM is provided in Supporting Information S2.

GRU is another modified version of RNN for time-series predictions (Cho et al., 2014; Chung et al., 2014). It simplifies the LSTM structure by combining the forget and input gates into a single update gate, and meanwhile merging the cell state and hidden state. GRU contains only two gates: the reset gate and the update gate. The reset gate controls how much of the past data should be forgotten, and how new input information is combined with the previous memory; the update gate regulates the balance between input and forgetting, thereby establishing long-term relationships for future prediction (see Supporting Information S3).

Transformer is a deep learning model relying on "self-attention" to compute representations of its input and output without using sequence-aligned RNNs (Vaswani et al., 2017). It consists of an encoder and a decoder with several Transformer blocks of the same architecture. The encoder generates encodings of inputs, and the decoder takes all the

encodings and uses their incorporated contextual information to generate the output sequence for future prediction (see Supporting Information S4).

The datasets for training and testing the three deep learning models are identical to those for the developed P-ED-LSTM model, with the data input of 12 online parameters (see Fig. 5a) and prediction output of WWTP effluent TN concentration at future time.

4.3. Data source

Data used to develop models for predicting effluent water quality were obtained from a WWTP in China, as seen in Supporting Information S1. Specifically, the bi-hourly data during the period of July 1, 2022 to June 30, 2023 were collected, matching the on-line data frequency of two hours (see Fig. S1). Statistics of WWTP influent and effluent data during this period are presented in Table S1.

Both Fig. S1 and Table S1 show that the recorded WWTP influent data varied considerably in terms of both water quantity and water quality. In more information, inflow rates of this WWTP fluctuated widely from 12.1 to 150 m^3/h with an average flow rate of 66.8 m^3/h and a coefficient of variation (CV) of 0.20. The CV of influent COD, NH_3-N, TP, and TN concentrations ranged between 0.30 and 0.53, demonstrating even higher values and associated variability compared to that of the inflow rates.

In response to influent characteristics, effluent TN concentration of the WWTP also varied substantially within the data period. Effluent TN fluctuated obviously between 1.44 and 25.0 mg/L, with a CV of 0.30. For this WWTP, the treated effluent should comply with China's Grade 1-A discharge standard for municipal WWTP (GB18918–2002). According to the standard, the discharge limit of effluent TN concentration is set to 15 mg/L. However, as seen in Table S1, there were observations exceeding this discharge limit, and the highest TN effluent concentration was up to 1.7 times the prescribed discharge limit due to sudden influent shocking load events.

4.4. Data processing

4.4.1. Data splitting

As aforementioned, the entire dataset was divided into two subsets, namely training and test datasets. Generally, the performance of machine learning models increases by feeding more data for the training stage than the test stage. Zhang et al. (2019) showed that a stronger capability and robustness in prediction and extrapolation could be offered when using 70~89% of the whole data for training. Zhu et al. (2023) suggested a data percentage cut with the proportion of test data ranging between 10%~40%. With these experiences, in our case, the first 80% of the whole time-series datasets were used for training, and the remaining 20% of the whole datasets were employed for testing purposes. In detail, for the data collected during the period of July 1, 2022 to June 30, 2023 (see Fig. S2), the first 3504 sets of data and the rest 876 sets of data were extracted for training and test of the deep learning models, respectively.

4.4.2. Data normalization

As the magnitude of online data varied obviously among influent, effluent, and treatment process monitoring parameters, the acquired online datasets were pre-normalized prior to their employment for model training and testing. The normalization was performed with the standard deviation normalization method that follows the equation below (Singh and Singh, 2020):

$$x' = \frac{x_i - \overline{x}}{x_{sd}} \tag{5}$$

where x' is the normalized data; x is the measured data; $\overline{x}$ is the mean value of the measured data, and x_{sd} is the standard deviation of the measured data.

4.5. Inspection index

4.5.1. Metric indicators

The metric indicators of MAPE, MAE, RMSE, and R^2 are calculated according to the mathematical expressions below (Tofallis, 2015):

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_{\mathrm{pred},i} - y_{\mathrm{meas},i}\right| \tag{6}$$

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_{\mathrm{pred},i} - y_{\mathrm{meas},i}}{y_{\mathrm{meas},i}}\right| \tag{7}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_{\mathrm{pred},i} - y_{\mathrm{meas},i}\right)^2} \tag{8}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_{\mathrm{pred},i} - y_{\mathrm{meas},i}\right)^2}{\sum_{i=1}^{n}\left(\overline{y}_{\mathrm{meas}} - y_{\mathrm{meas},i}\right)^2} \tag{9}$$

where n is the number of predicted time-series data in the training or test dataset; $y_{\mathrm{pred},i}$ is the predicted time series of the WWTP effluent quality data (effluent TN concentration in this case), and $y_{\mathrm{meas},i}$ is the measured time series of the WWTP effluent quality data; $\overline{y}_{\mathrm{meas}}$ is the average of y_{meas} over the n data.

The larger the R^2 and the smaller the MAPE, MAE, and RMSE, the closer the predicted value is to the measured value. A value of 1 for R^2 and 0 for MAPE, MAE, and RMSE indicates the highest accuracy. In this way, these metrics offer a quantitative measurement of the models' accuracy to determine the deep learning model with the best performance.

In addition, the early-warning accuracy is introduced to measure the model's reliability for successful predictions of alert and no alert cases under ahead time, which is quantified using the following formula (Liu et al., 2022):

$$\mathrm{AC} = \frac{t_p + t_m}{n} \tag{10}$$

where AC is the early-warning accuracy; t_p is the number of successful predictions for no alert cases which is below the discharge limit; t_m is the number of successful predictions for alert cases which is above the discharge limit; n is the number of total cases.

The bigger the AC in this equation, the higher the model's reliability is for early-warning of the cases exceeding the discharge limit.

4.5.2. The Taylor diagram

The Taylor diagram is a method that provides an overall measurement of the performance of the deep learning models (Zounemat-Kermani et al., 2021). It is characterized by a combination of three metrics including correlation coefficient, standard deviation, and CRMSE, which are determined following the mathematical expressions below:

$$\sigma_{\mathrm{meas}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\overline{y}_{\mathrm{meas}} - y_{\mathrm{meas},i}\right)^2} \tag{11}$$

$$\sigma_{\mathrm{pred}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\overline{y}_{\mathrm{pred}} - y_{\mathrm{pred},i}\right)^2} \tag{12}$$

$$r = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(\overline{y}_{\mathrm{meas}} - y_{\mathrm{meas},i}\right)\left(\overline{y}_{\mathrm{pred}} - y_{\mathrm{pred},i}\right)}{\sigma_{\mathrm{meas}}\sigma_{\mathrm{pred}}} \tag{13}$$

$$\mathrm{CRMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[\left(\bar{y}_{\mathrm{meas}} - y_{\mathrm{meas},i}\right)\left(\bar{y}_{\mathrm{pred}} - y_{\mathrm{pred},i}\right)\right]^2} \quad (14)$$

where σ_{meas} and σ_{pred} are the standard deviation of measured and predicted time-series values, respectively; r is the correlation coefficient between measured and predicted values; $\bar{y}_{\mathrm{pred}}$ is the average of predicted time series y_{pred} over the n data.

With the calculation of the three metrics, the overall performance of deep learning models is generalized as one definite point (hereinafter referred to as "model point") in the Taylor diagram. Meanwhile, there is also a generalized point representing the measurement data series for effluent water quality (hereinafter referred to as "measurement point"). Hence, the distance between the model point and measurement point quantifies the performance of the employed deep learning model. The closer the model point is to the measurement point, the better the model's overall performance is.

4.5.3. Model interpretation

The Shapley additive explanation (SHAP) was also employed to further interpret the behavior of the developed machine learning model (Lundberg and Lee, 2017; Zhang et al., 2023). It is based on the Shapley value derived from cooperative game theory, which assigns each feature (i.e., input parameter) a value to indicate its contribution to the model output (Li et al., 2022; Wang et al., 2022). The larger the absolute Shapley value is, the higher the contribution of the feature is. Therefore, an advantage of the SHAP value is its ability to describe how the model prediction varies based on changes in the values of the model inputs. Mathematically, the Shapley value is defined as

$$g(z') = \varphi_0 + \sum_{m=1}^{M}\varphi_m z'_m \quad (15)$$

where $g(z')$ is the explanation model; φ_0 is the expected value of model prediction, which is equal to mean model prediction at the dataset; M is the number of all input parameters; φ_m is the Shapley value of input parameter m; $z' \in \{0,1\}$ is the simplified features with 1 or 0 indicating whether a model feature is present or absent when computing the Shapley value.

CRediT authorship contribution statement

Hailong Yin: Writing – review & editing, Writing – original draft, Methodology, Conceptualization. **Yongqi Chen:** Writing – original draft, Formal analysis, Conceptualization. **Jingshu Zhou:** Investigation, Data curation. **Yifan Xie:** Formal analysis. **Qing Wei:** Visualization. **Zuxin Xu:** Writing – review & editing, Supervision.

Declaration of competing interest

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

Acknowledgements

This work was financially supported by the National Natural Science Foundation of China (Grant No. 52170103), the National Key R&D Program of China (Grant No. 2021YFC3200703).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.wroa.2024.100291.

Data availability

Data will be made available on request.

References

Alvi, M., Batstone, D., Mbamba, C., Keymer, P., French, T., Ward, A., Dwyer, J., Cardell-Oliver, R., 2023. Deep learning in wastewater treatment: a critical review. Water. Res. 245. https://doi.org/10.1016/j.watres.2023.120518.

Badran, I., Manasrah, A.D., Nassar, N.N., 2019. A combined experimental and density functional theory study of metformin oxy-cracking for pharmaceutical wastewater treatment. RSC. Adv. 9, 13403–13413. https://doi.org/10.1039/c9ra01641d.

Chang, P., Zhang, S., Wang, Z., 2024. Soft sensor of the key effluent index in the municipal wastewater treatment process based on Transformer. IEEE Trans. Ind. Inform. 20, 4021–4028. https://doi.org/10.1109/TII.2023.3316179.

Cheng, T., Harrou, F., Kadri, F., Sun, Y., Leiknes, T., 2020. Forecasting of wastewater treatment plant key features using deep learning-based models: A case study. IEEe Access. 8, 184475–184485. https://doi.org/10.1109/ACCESS.2020.3030820.

Ching, P.M.L., Zou, X., Wu, D., So, R.H.Y., Chen, G.H., 2022. Development of a wide-range soft sensor for predicting wastewater BOD5 using an extreme gradient boosting (XGBoost) machine. Environ. Res. 210. https://doi.org/10.1016/j.envres.2022.112953.

Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y., 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078.

Chung, J., Gulcehre, C., Cho, K., Bengio, Y., 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555.

Cimen Mesutoglu, O., Gok, O., 2024. Prediction of COD in industrial wastewater treatment plant using an artificial neural network. Sci. Rep. 14. https://doi.org/10.1038/s41598-024-64634-z, 13750–13750.

Guo, H., Jeong, K., Lim, J., Jo, J., Kim, Y.M., Park, J.-P., Kim, J.H., Cho, K.H., 2015. Prediction of effluent concentration in a wastewater treatment plant using machine learning models. J. Environ. Sci. 32, 90–101. https://doi.org/10.1016/j.jes.2015.01.007.

Hou, Q., Gao, Z., Duan, Z., Yu, M., 2024. A Study on the Transformer-CNN imputation method for turbulent heat flux dataset in the Qinghai-Tibet Plateau grassland. EGUsphere 2024, 1–24. https://doi.org/10.5194/egusphere-2023-2685.

Hu, T., Zhang, Y., Wang, X., Sha, J., Dai, H., Xiong, Z., Wang, D., Zhang, F., Liu, H., 2024. Optimized convolutional neural networks for fault diagnosis in wastewater treatment processes. Environ. Sci.-Water Res. Technol. 10, 364–375. https://doi.org/10.1039/d3ew00619k.

Huang, R., Ma, C., Ma, J., Huangfu, X., He, Q., 2021. Machine learning in natural and engineered water systems. Water. Res. 205. https://doi.org/10.1016/j.watres.2021.117666.

Huang, Y., Xie, Y., Wu, Y., Meng, F., He, C., Zou, H., Wang, X., Shui, A., Liu, S., 2023. Modeling indirect greenhouse gas emissions sources from urban wastewater treatment plants: Integrating machine learning models to compensate for sparse parameters with abundant observations. Environ. Sci. Technol. 57, 19860–19870. https://doi.org/10.1021/acs.est.3c06482.

Jaffari, Z., Na, S., Abbas, A., Park, K., Cho, K., 2024. Digital imaging-in-flow (FlowCAM) and probabilistic machine learning to assess the sonolytic disinfection of cyanobacteria in sewage wastewater. J. Hazard. Mater. 468. https://doi.org/10.1016/j.jhazmat.2024.133762.

Jiang, Y., Li, C., Zhang, Y., Zhao, R., Yan, K., Wang, W., 2021. Data-driven method based on deep learning algorithm for detecting fat, oil, and grease (FOG) of sewer networks in urban commercial areas. Water. Res. 207. https://doi.org/10.1016/j.watres.2021.117797.

Kao, I., Zhou, Y., Chang, L., Chang, F., 2020. Exploring a long short-term memory based encoder-decoder framework for multi-step-ahead flood forecasting. J. Hydrol. 583. https://doi.org/10.1016/j.jhydrol.2020.124631.

Karamnia, B., Roudsari, F.P., Mehrnia, M.R., 2024. Unsteady state municipal wastewater treatment: MBR response to organic and hydraulic shock-loads. Urban. Water. J. https://doi.org/10.1080/1573062X.2024.2312513.

Kim, H., Hong, S., Jeong, K., Kim, D., Joo, G., 2019. Determination of sensitive variables regardless of hydrological alteration in artificial neural network model of chlorophyll a: case study of Nakdong River. Ecol. Model. 398, 67–76. https://doi.org/10.1016/j.ecolmodel.2019.02.003.

Li, G., Cui, Q., Wei, S., Wang, X., Xu, L., He, L., Kwong, T.C.H., Tang, Y., 2023a. Long short-term memory network-based wastewater quality prediction model with sparrow search algorithm. Int. J. Wavelets Multiresolution Inf. Process. https://doi.org/10.1142/S0219691323500194.

Li, J., Dong, J., Chen, Z., Li, X., Yi, X., Niu, G., He, J., Lu, S., Ke, Y., Huang, M., 2023b. Free nitrous acid prediction in ANAMMOX process using hybrid deep neural network model. J. Environ. Manage. 345. https://doi.org/10.1016/j.jenvman.2023.118566.

Li, J., Du, Z., Liu, J., Xu, L., He, L., Gu, L., Cheng, H., He, Q., 2024. Analysis of factors influencing the energy efficiency in Chinese wastewater treatment plants through machine learning and SHapley additive explanations. Sci. Total Environ. 920. https://doi.org/10.1016/j.scitotenv.2024.171033.

Li, L., Qiao, J., Yu, G., Wang, L., Li, H.-Y., Liao, C., Zhu, Z., 2022. Interpretable tree-based ensemble model for predicting beach water quality. Water. Res. 211. https://doi.org/10.1016/j.watres.2022.118078.

Liu, M., He, J., Huang, Y., Tang, T., Hu, J., Xiao, X., 2022. Algal bloom forecasting with time-frequency analysis: a hybrid deep learning approach. Water. Res. 219. https://doi.org/10.1016/j.watres.2022.118591.

Liu, W., Liu, T., Liu, Z., Luo, H., Pei, H., 2023. A novel deep learning ensemble model based on two-stage feature selection and intelligent optimization for water quality prediction. Environ. Res. 224. https://doi.org/10.1016/j.envres.2023.115560.

Liu, X., Shi, Q., Liu, Z., Yuan, J., 2021. Using LSTM neural network based on improved PSO and attention mechanism for predicting the effluent COD in a wastewater treatment plant. IEEe Access. 9, 146082–146096. https://doi.org/10.1109/ACCESS.2021.3123225.

Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. A unified approach to interpreting model predictions. Adv. Neural Inf. Process. Syst. 30.

Lv, J., Du, L., Lin, H., Wang, B., Yin, W., Song, Y., Chen, J., Yang, J., Wang, A., Wang, H., 2024. Enhancing effluent quality prediction in wastewater treatment plants through the integration of factor analysis and machine learning. Bioresour. Technol. 393. https://doi.org/10.1016/j.biortech.2023.130008.

Manav-Demir, N., Gelgor, H.B., Oz, E., Ilhan, F., Ulucan-Altuntas, K., Tiwary, A., Debik, E., 2024. Effluent parameters prediction of a biological nutrient removal (BNR) process using different machine learning methods: a case study. J. Environ. Manage. 351. https://doi.org/10.1016/j.jenvman.2023.119899.

Maurya, A.K., Reddy, B.S., Theerthagiri, J., Narayana, P.L., Park, C.H., Hong, J.K., Yeom, J.-T., Cho, K.K., Reddy, N.S., 2021. Modeling and optimization of process parameters of biofilm reactor for wastewater treatment. Sci. Total Environ. 787. https://doi.org/10.1016/j.scitotenv.2021.147624.

Mjalli, F.S., Al-Asheh, S., Alfadala, H.E., 2007. Use of artificial neural network black-box modeling for the prediction of wastewater treatment plants performance. J. Environ. Manage. 83, 329–338. https://doi.org/10.1016/j.jenvman.2006.03.004.

Nourani, V., Asghari, P., Sharghi, E., 2021. Artificial intelligence based ensemble modeling of wastewater treatment plant using jittered data. J. Clean. Prod. 291. https://doi.org/10.1016/j.jclepro.2020.125772.

Ocampo-Perez, R., Rivera-Utrilla, J., Mota, A.J., Sanchez-Polo, M., Leyva-Ramos, R., 2016. Effect of radical peroxide promoters on the photodegradation of cytarabine antineoplastic in water. Chem. Eng. J. 284, 995–1002. https://doi.org/10.1016/j.cej.2015.08.162.

Sadri Moghaddam, S., Mesghali, H., 2023. A new hybrid ensemble approach for the prediction of effluent total nitrogen from a full-scale wastewater treatment plant using a combined trickling filter-activated sludge system. Environ. Sci. Pollut. Res. 30, 1622–1639. https://doi.org/10.1007/s11356-022-21864-w.

Sharafati, A., Asadollah, S.B.H.S., Hosseinzadeh, M., 2020. The potential of new ensemble machine learning models for effluent quality parameters prediction and related uncertainty. Process Saf. Environ. Prot. 140, 68–78. https://doi.org/10.1016/j.psep.2020.04.045.

Singh, D., Singh, B., 2020. Investigating the impact of data normalization on classification performance. Appl. Soft Comput. 97. https://doi.org/10.1016/j.asoc.2019.105524.

Suresh, V., Aksan, F., Janik, P., Sikorski, T., Revathi, B., 2022. Probabilistic LSTM-Autoencoder based hour-ahead solar power forecasting model for intra-day electricity market participation: a Polish case study. IEEe Access. 10, 110628–110638. https://doi.org/10.1109/ACCESS.2022.3215080.

Suresh, K., Tang, T., van Vliet, M.T.H., Bierkens, M.F.P., Strokal, M., Sorger-Domenigg, F., Wada, Y., 2023. Recent advancement in water quality indicators for eutrophication in global freshwater lakes. Environ. Res. Lett. 18. https://doi.org/10.1088/1748-9326/acd071.

Tofallis, C., 2015. A better measure of relative prediction accuracy for model selection and model estimation. J. Oper. Res. Soc. 66, 1352–1362. https://doi.org/10.1057/jors.2014.103.

Van Houdt, G., Mosquera, C., Napoles, G., 2020. A review on the long short-term memory model. Artif. Intell. Rev. 53, 5929–5955. https://doi.org/10.1007/s10462-020-09838-1.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. Adv. Neural Inf. Process. Syst. 30.

Wang, D., Thunell, S., Lindberg, U., Jiang, L., Trygg, J., Tysklind, M., 2022. Towards better process management in wastewater treatment plants: process analytics based on SHAP values for tree-based machine learning methods. J. Environ. Manage. 301. https://doi.org/10.1016/j.jenvman.2021.113941.

Wang, X., Ratnaweera, H., Holm, J.A., Olsbu, V., 2017. Statistical monitoring and dynamic simulation of a wastewater treatment plant: a combined approach to achieve model predictive control. J. Environ. Manage. 193, 1–7. https://doi.org/10.1016/j.jenvman.2017.01.079.

Wang, Z., Man, Y., Hu, Y., Li, J., Hong, M., Cui, P., 2019. A deep learning based dynamic COD prediction model for urban sewage. Environ. Sci.-Water Res. Technol. 5, 2210–2218. https://doi.org/10.1039/c9ew00505f.

Wang, Z., Su, X., Ding, Z., 2021. Long-term traffic prediction based on LSTM encoder-decoder architecture. IEEE Trans. Intell. Transp. Syst. 22, 6561–6571. https://doi.org/10.1109/TITS.2020.2995546.

Xie, Y., Chen, Y., Lian, Q., Yin, H., Peng, J., Sheng, M., Wang, Y., 2022. Enhancing real-time prediction of effluent water quality of wastewater treatment plant based on improved feedforward neural network coupled with optimization algorithm. Water. (Basel) 1053. https://doi.org/10.3390/w14071053, 3.5299 14.

Xie, Y., Chen, Y., Wei, Q., Yin, H., 2024. A hybrid deep learning approach to improve real-time effluent quality prediction in wastewater treatment plant. Water. Res. 250, 121092. https://doi.org/10.1016/j.watres.2023.121092.

Xu, R., Wang, Y., Chen, Z., 2023. A hybrid approach to predict battery health combined with attention-based transformer and online correction. J. Energy Storage 65. https://doi.org/10.1016/j.est.2023.107365.

Yan, K., Li, C., Zhao, R., Zhang, Y., Duan, H., Wang, W., 2023. Predicting the ammonia nitrogen of wastewater treatment plant influent via integrated model based on rolling decomposition method and deep learning algorithm. Sustain. Cities Soc. 94. https://doi.org/10.1016/j.scs.2023.104541.

Yang, Y., Kim, K.-R., Kou, R., Li, Y., Fu, J., Zhao, L., Liu, H., 2022. Prediction of effluent quality in a wastewater treatment plant by dynamic neural network modeling. Process Saf. Environ. Prot. 158, 515–524. https://doi.org/10.1016/j.psep.2021.12.034.

Zeng, A., Chen, M., Zhang, L., Xu, Q., 2023. Are Transformers effective for time series forecasting? Proc. AAAI Conf. Artif. Intell. 37, 11121–11128. https://doi.org/10.1609/aaai.v37i9.26317.

Zhang, L., Liu, Z., Xiao, W., Meng, D., 2022. Probabilistic vehicle trajectory prediction based on LSTM encoder-decoder and attention mechanism. SAE Technical Paper. https://doi.org/10.4271/2022-01-7106, 2022-01-7106.

Zhang, Y., Gao, X., Smith, K., Inial, G., Liu, S., Conil, L., Pan, B., 2019. Integrating water quality and operation into prediction of water production in drinking water treatment plants by genetic algorithm enhanced artificial neural network. Water. Res. 164. https://doi.org/10.1016/j.watres.2019.114888.

Zhang, Y., Li, C., Duan, H., Yan, K., Wang, J., Wang, W., 2023. Deep learning based data-driven model for detecting time-delay water quality indicators of wastewater treatment plant influent. Chem. Eng. J. 467. https://doi.org/10.1016/j.cej.2023.143483.

Zheng, Z., Zhang, Z., 2024. A stochastic recurrent encoder decoder network for multistep probabilistic wind power predictions. IEEE Trans. Neural Netw. Learn. Syst. 35, 9565–9578. https://doi.org/10.1109/TNNLS.2023.3234130.

Zhong, S., Zhang, K., Bagheri, M., Burken, J., Gu, A., Li, B., Ma, X., Marrone, B., Ren, Z., Schrier, J., Shi, W., Tan, H., Wang, T., Wang, X., Wong, B., Xiao, X., Yu, X., Zhu, J., Zhang, H., 2021. Machine learning: new ideas and tools in environmental science and engineering. Environ. Sci. Technol. 55, 12741–12754. https://doi.org/10.1021/acs.est.1c01339.

Zhu, J., Yang, M., Ren, Z., 2023. Machine learning in environmental research: common pitfalls and best practices. Environ. Sci. Technol. https://doi.org/10.1021/acs.est.3c00026.

Zhu, M., Wang, J., Yang, X., Zhang, Y., Zhang, L., Ren, H., Wu, B., Ye, L., 2022. A review of the application of machine learning in water quality evaluation. Eco-Environ. Health Online 1, 107–116. https://doi.org/10.1016/j.eehl.2022.06.001.

Zounemat-Kermani, M., Mahdavi-Meymand, A., Hinkelmann, R., 2021. A comprehensive survey on conventional and modern neural networks: application to river flow forecasting. Earth Sci. Inform. 14, 893–911. https://doi.org/10.1007/s12145-021-00599-1.