

Article

A Multivariable Probability Density-Based Auto-Reconstruction Bi-LSTM Soft Sensor for Predicting Effluent BOD in Wastewater Treatment Plants

Wenting Li, Yonggang Li, Dong Li  and Jiayi Zhou

School of Automation, Central South University, Changsha 410083, China; liwenting@csu.edu.cn (W.L.)

* Correspondence: lddscut@163.com

Abstract: The precise detection of effluent biological oxygen demand (BOD) is crucial for the stable operation of wastewater treatment plants (WWTPs). However, existing detection methods struggle to meet the evolving drainage standards and management requirements. To address this issue, this paper proposed a multivariable probability density-based auto-reconstruction bidirectional long short-term memory (MPDAR-Bi-LSTM) soft sensor for predicting effluent BOD, enhancing the prediction accuracy and efficiency. Firstly, the selection of appropriate auxiliary variables for soft-sensor modeling is determined through the calculation of k-nearest-neighbor mutual information (KNN-MI) values between the global process variables and effluent BOD. Subsequently, considering the existence of strong interactions among different reaction tanks, a Bi-LSTM neural network prediction model is constructed with historical data. Then, a multivariate probability density-based auto-reconstruction (MPDAR) strategy is developed for adaptive updating of the prediction model, thereby enhancing its robustness. Finally, the effectiveness of the proposed soft sensor is demonstrated through experiments using the dataset from Benchmark Simulation Model No.1 (BSM1). The experimental results indicate that the proposed soft sensor not only outperforms some traditional models in terms of prediction performance but also excels in avoiding ineffective model reconstructions in scenarios involving complex dynamic wastewater treatment conditions.

Keywords: effluent BOD; soft sensor; Bi-LSTM; MPDAR strategy; wastewater treatment



Citation: Li, W.; Li, Y.; Li, D.; Zhou, J. A Multivariable Probability Density-Based Auto-Reconstruction Bi-LSTM Soft Sensor for Predicting Effluent BOD in Wastewater Treatment Plants. *Sensors* **2024**, *24*, 7508. <https://doi.org/10.3390/s24237508>

Academic Editor: Chase Wu

Received: 6 October 2024

Revised: 22 November 2024

Accepted: 23 November 2024

Published: 25 November 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the rapid increase in population and the development of urbanization, more and more urban wastewater needs to be treated urgently [1,2]. The activated sludge mechanism, recognized as a convenient and efficient treatment technology, has seen widespread adoption in urban wastewater treatment plants in recent years [3]. Effluent biological oxygen demand (BOD) serves as a vital parameter for assessing discharge standards and indirectly reflects the amount of biodegradable organic matter in wastewater [4,5]. Therefore, timely and accurate prediction of effluent BOD is crucial as it can provide valuable insights for the safe operation and scientific management of wastewater treatment plants. At present, the commonly used detection methods encompass offline sampling and testing, online instrument monitoring, and the utilization of soft sensor technology [6]. Among these, soft sensor technology has garnered significant attention in recent years due to its convenient operation, low economic cost, and precise detection outcomes [7–9]. Notably, the development of storage and computational capabilities of hardware devices has propelled the use of data-driven soft sensors as a prominent detection method. Meng et al. proposed an adaptive task-oriented radial basis function (ATO-RBF) network to predict the effluent BOD through pruning the RBF nodes without sacrificing the learning accuracy [10]. To improve the prediction efficiency of effluent BOD, Pattnaik et al. proposed a cloud and edge layer collaborative soft sensor system architecture [11]. However, there are numerous variables in wastewater treatment processes, and the operational conditions constantly

change due to internal and external factors; thus, how to improve the robustness of models while reducing time consumption of training processes is a worthwhile research topic.

For one thing, variable selection is a crucial step in the implementation of soft sensor technology, aimed at diminishing the dimensionality of sample data and simplifying model complexity [12]. The conventional methods for variable selection typically depend on mechanistic analysis and manual expertise to identify relevant process variables associated with BOD. However, the lack of a solid theoretical foundation often leads to imprecise selection. Then, principal component analysis (PCA) offers a means of dimensionality reduction through statistical examination of raw data, significantly easing the computational load in modeling [13]. However, PCA falls under unsupervised learning, resulting in reduced interpretability of the model due to the dimensionality-reduced variables that are not the original process variables. To address this problem, some researchers explored selecting the auxiliary variables depending on the Pearson correlation coefficient (PCC) [14] and mutual information (MI) [15]. Nonetheless, these approaches fail to consider spatial characteristics among variables in wastewater treatment processes and the constantly increasing computational burden.

For another, the utilization of a dependable prediction model is vital for enhancing the precision of soft sensor detection outcomes [16]. Presently, advanced deep learning algorithms like convolutional neural networks (CNNs) [17], graph neural networks (GNNs) [18], and recurrent neural networks (RNNs) [19] have gained substantial traction in wastewater treatment processes of variable prediction owing to their exceptional capacity for fitting complex data. Regrettably, the lengthy training duration of deep neural networks often relegates such soft sensors to offline modeling but online application [20]. At the same time, instances of abrupt alterations in external or internal conditions can severely compromise the robustness of the model, causing a significant decline in predictive accuracy. To solve this problem, scholars have directed their focus towards model correction and reconstruction [21–23]. For instance, Liu et al. proposed a method of structural pruning, involving iterative adjustments to the model architecture to accommodate dynamic condition changes [24]. Nevertheless, corrective measures tend to fall short when confronted with extreme scenarios. Therefore, Li et al. proposed the use of a moving window (MW) to continually update training data and reconstruct the prediction model [25]. However, the iterative reconstruction process demands extensive time for parameter fine-tuning and optimization, thereby impeding the timeliness of prediction results.

In response to the above issues, this paper proposed a multivariable probability density-based auto-reconstruction Bi-LSTM (MPDAR-Bi-LSTM) soft sensor for predicting effluent BOD in wastewater treatment plants. The main contributions and novelties of this study are as follows:

- (1) To select auxiliary variables closely related to effluent BOD from global process variables for subsequent modeling, the k-nearest-neighbor (KNN) method is used instead of the original probability density computation method. This enhances the computational efficiency of mutual information (MI) values as well as incorporates an exploration of the spatial characteristic among process variables.
- (2) Considering that each process variable in WWTP is affected by its own temporal changes, a prediction model is constructed leveraging the Bi-LSTM neural network, which consists of forward and backward LSTM layers to extract the latent information during data transmission.
- (3) To enhance the robustness of the prediction model while ensuring its prediction efficiency, a multivariable probability density-based auto-reconstruction (MPDAR) strategy is proposed. It determines the necessity of model reconstruction by comparing the change rate of multivariate mathematical expectations, striking a balance between the stability and timeliness of the prediction model.

The rest of this paper is organized in accordance with the following scheme. In Section 2, we introduce the related works of KNN-MI for variable selection, Bi-LSTM for modeling prediction, and MPDAR strategy for model updating, which are the foundations

of the proposed MPDAR-Bi-LSTM soft sensor. In Section 3, the proposed soft sensor is experimentally validated and analyzed on BSM1. Finally, the conclusions are given in Section 4, and some potential future research directions are provided.

2. Methodology

The main content of Section 2 includes the following four points. Section 2.1 provides a KNN-MI method to select auxiliary variables closely related to effluent BOD from global process variables for subsequent modeling, which enhances the computational efficiency of MI values as well as incorporates an exploration of the spatial characteristics among process variables. Section 2.2 introduces the theoretical knowledge of the Bi-LSTM neural network, which serves as the foundation for subsequent prediction models. Section 2.3 provides an MPDAR strategy that determines the necessity of prediction model reconstruction, which is the key to enhancing the robustness of the prediction model while ensuring its prediction efficiency. On the basis of all the theoretical knowledge and innovative methods mentioned above, the MPDAR-Bi-LSTM soft sensor for predicting effluent BOD in wastewater treatment plants is proposed. And Section 2.4 provides the specific steps of the proposed MPDAR-Bi-LSTM soft sensor depending on the previous statements.

2.1. KNN-MI for Variable Selection

Mutual information (MI) is a measure of the degree of interdependence between two variables [26]. The larger the MI value, the stronger both the linear and nonlinear correlations between two variables. Currently, it has been applied in various fields such as data science, information security, and process industry. According to its definition, the specific calculation formula is as follows:

$$I(X, Y) = \iint \rho(x, y) \lg \frac{\rho(x, y)}{\rho(x)\rho(y)} dx dy | x \in X, y \in Y \quad (1)$$

where $\rho(x, y)$ represents the joint probability density value between two variables x and y . $\rho(x)$ and $\rho(y)$ are the marginal probability density values of variables x and y , respectively. $I(X, Y)$ is the MI value between vectors X and Y .

In order to calculate the MI values more conveniently, this paper uses KNN to directly approximate their probability density values [27]. The main idea is to first identify k -nearest-neighbor samples, then find the number of other samples whose distances from the k -nearest-neighbor samples are less than the maximum distance between the current sample and the k -nearest-neighbor samples, and finally obtain the MI values through counting the numbers. The specific calculation formula is as follows:

$$I(X, Y) = \Psi(x) - 1/k - \langle \Psi(n_x) + \Psi(n_y) \rangle + \Psi(S) | x \in X, y \in Y \quad (2)$$

where $\Psi(\cdot)$ represents the digamma function. k is the number of nearest-neighbor samples, usually between 2 and 6. $\varepsilon_i/2$ is defined as the distance from the i -th sample x_i or y_i to the k -th-nearest-neighbor sample, so n_x is the number of samples with a distance less than $\varepsilon_i/2$ from x_i ; n_y is the number of samples with a distance less than $\varepsilon_i/2$ from y_i . $\langle \cdot \rangle$ is the average value of the digamma function for all variables, $\langle \cdot \rangle = \sum_{i=1}^N E(\cdot)/S$; S represents the number of samples.

KNN-MI can more conveniently identify the correlation between the target variable and other process variables. When given a suitable threshold λ , we can directly select the most appropriate variables for subsequent modeling.

2.2. Bi-LSTM for Modeling Prediction

Deep learning algorithms are an emerging method for solving the problem of variable prediction in modern industries. Notably, recurrent neural networks like LSTM stand out among other neural networks due to their distinctive “memory function” [28]. LSTM is usually composed of a forget gate, an input gate, and an output gate. The structural

diagram is shown in Figure 1a. These components serve distinct purposes: the forget gate manages the retention or deletion of existing information, the input gate deals with receiving and storing new data, while the output gate regulates the output contributions of individual units.

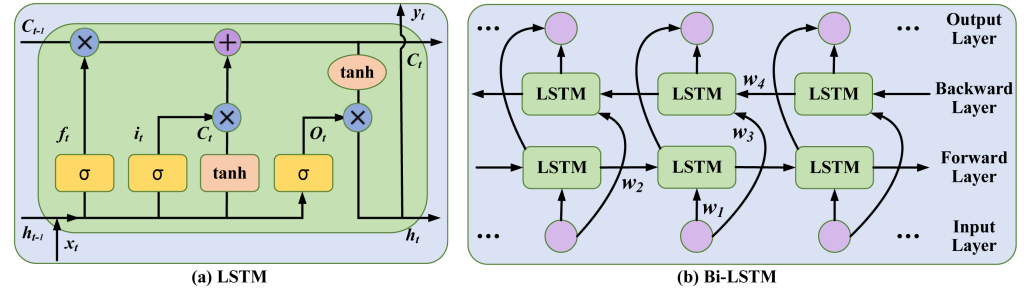


Figure 1. The structural diagrams of LSTM and Bi-LSTM.

In the forget gate, the process of information transmission is as follows:

$$f_t = \sigma(\omega_{f_h}[h_{t-1}], \omega_{f_x}[x_t], b_f) \quad (3)$$

where f_t represents the output of this gate, between 0 and 1. 0 indicates completely removing the value, and 1 implies reserving the whole value. σ is a sigmoid activation function. ω_{f_h} and ω_{f_x} are the weight matrices; b_f is the bias vector of the forget gate. h_{t-1} and x_t are the output of the last moment and the input of the current time, respectively.

In the input gate, this gate consists of two layers: (1) a sigmoid layer, and (2) a “tanh” layer. The process of information transmission is as follows:

$$\begin{aligned} i_t &= \sigma(\omega_{i_h}[h_{t-1}], \omega_{i_x}[x_t], b_i) \\ \tilde{C}_t &= \tanh(\omega_{c_h}[h_{t-1}], \omega_{c_x}[x_t], b_c) \end{aligned} \quad (4)$$

where i_t is between 0 and 1, which represents the transmission proportion of $\tilde{C}_t$. $\tilde{C}_t$ is a newly added candidate status during the current moment. $\tanh$ is a commonly used activation function. ω_{i_h} and ω_{i_x} the weight matrices. b_i and b_c are the bias vectors of the input gate.

Afterwards, the current state C_t is updated through the forget gate and input gate, the mathematical equation is as follows:

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \quad (5)$$

where f_t represents the output of the forget gate, i_t represents the transmission proportion of $\tilde{C}_t$, and C_{t-1} represents the previous state.

In the output gate, the process of information transmission is as follows:

$$\begin{aligned} O_t &= \sigma(\omega_{o_h}[h_{t-1}], \omega_{o_x}[x_t], b_o) \\ h_t &= O_t \times \tanh(C_t) \end{aligned} \quad (6)$$

where O_t is the output of the output gate, between 0 and 1. h_t represents the final output value. ω_{o_h} and ω_{o_x} are the weight matrices, and b_o is the bias vector of the forget gate.

Bi-LSTM is an extension of LSTM, which consists of two layers of LSTM networks: one for forward propagation and the other for backward propagation [29]. Two layers can improve learning long-term dependencies. The structural diagram of Bi-LSTM is shown in Figure 1b.

In the forward propagation layer, the process of information transmission is as follows:

$$L_t^f = \sigma(\omega_1[x_t], \omega_2[L_{t-1}^f], b_{L^f}) \quad (7)$$

In the backward propagation layer, the process of information transmission is as follows:

$$L_t^b = \sigma(\omega_3[x_t], \omega_4[L_{t-1}^b], b_{L^b}) \quad (8)$$

where L_t^f and L_t^b are the outputs of the forward propagation layer and backward propagation layer, respectively. $\omega_1, \omega_2, \omega_3$, and ω_4 are the weight matrices. b_{L^f} and b_{L^b} are bias vectors.

In this paper, we constructed a Bi-LSTM-based soft sensor for variable prediction. Benefiting from the bidirectional information transmission process, Bi-LSTM is capable of revealing latent information and simulating the wastewater treatment process, consequently enhancing the prediction performance of the model.

2.3. MPDAR Strategy for Model Updating

Probability density is a statistical information measurement that serves as a valuable tool for characterizing the distribution properties of sample data. For the convenience of solving, we choose kernel density estimation (KDE) to calculate the probability density value of each process variable [30]. The KDE formula is as follows:

$$p(x) = \frac{1}{nh} \sum_{i=1}^n \phi\left(-\frac{x_i - x}{h}\right) \quad (9)$$

where $p(\cdot)$ represents the probability density value. n is the number of output data. h is a hyperparameter named window width. $\phi(\cdot)$ is a Gaussian kernel function, defined as follows:

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\|x\|^2}{2}} \quad (10)$$

Then, we calculate the mathematical expectation of vector X , denoted as $E(X)$. The formula for calculating $E(X)$ is as follows:

$$E(X) = \int_{-\infty}^{\infty} x \cdot p(x) dx \quad (11)$$

Finally, the change rate of $E(X)$ is calculated as follows:

$$J = \left| \frac{E(X_t) - E(X_{t-1})}{E(X_{t-1})} \right| \times 100\% \quad (12)$$

We extend Equation (11) to multivariate systems, and the calculation formula of J is as follows:

$$J = \left| \frac{\sum_{i=1}^m W_i \cdot ((E_i(X_t) - E_i(X_{t-1})))}{\sum_{i=1}^m (W_i E_i(X_{t-1}))} \right| \times 100\% \quad (13)$$

where W_i represents the weight value, which is determined by relying on the previous KNN-MI values. The specific calculation formula for W_i is as follows:

$$W_i = \frac{I(X_i, Y)}{\sum_{i=1}^m I(X_i, Y)} \quad (14)$$

where $I(X_i, Y)$ is the KNN-MI value between X_i and Y . m is the number of auxiliary variables.

Comparing with the provided threshold, if J surpasses the designated threshold, it signifies a noteworthy alteration in the distribution attributes of the training data, necessitating a reconstruction of the model. Conversely, if J falls below the threshold, there is no imperative for model reconstruction. Therefore, selective reconstruction of the model can be applied instead of blind reconstruction.

2.4. The Proposed MPDAR-Bi-LSTM Soft Sensor

This subsection provides the specific steps of the proposed MPDAR-Bi-LSTM soft sensor depending on the previous statements. Figure 2 shows the algorithm flowchart of the proposed soft sensor. The outlined steps are as follows.

Step 1: Collect sample data from the target WWTP and perform preprocessing of the collected data.

Step 2: Calculate the KNN-MI values between each process variable and effluent BOD, and then select the auxiliary variables for subsequent modeling according to the given threshold λ .

Step 3: Construct a Bi-LSTM-based prediction model using the historical training data.

Step 4: Calculate the change rate of mathematical expectation of the training dataset before and after updates, and determine whether to reconstruct the prediction model.

Step 5: Use the final model to predict the effluent BOD value of the test data, and assess the accuracy through comparing with the real values.

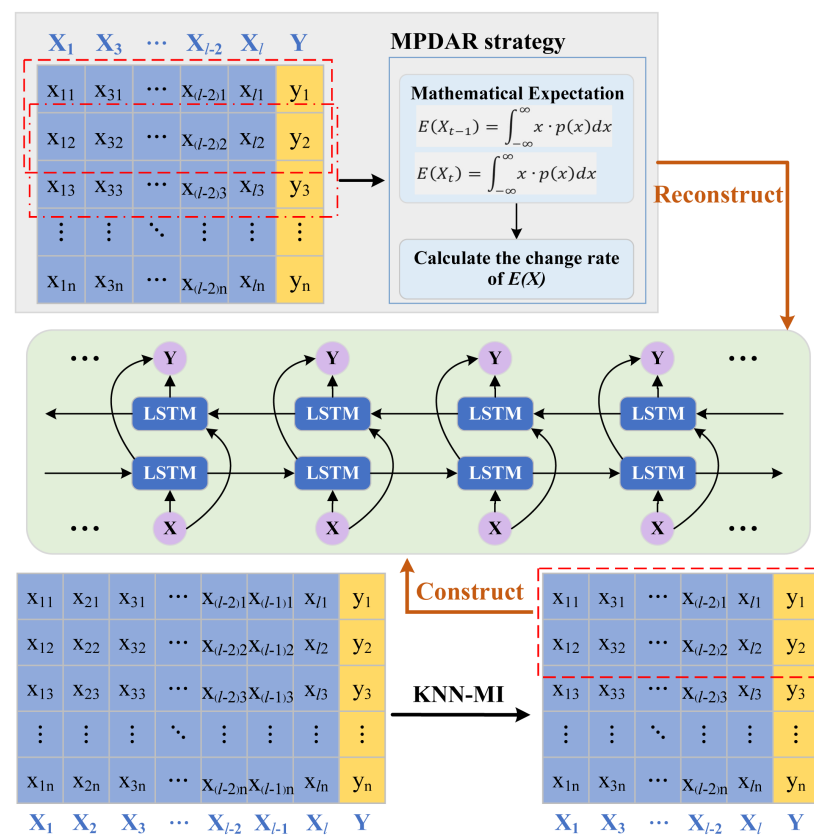


Figure 2. The algorithm flowchart of the proposed MPDAR-Bi-LSTM soft sensor.

3. Experiments

3.1. Background Description

BSM1 is a universal wastewater treatment simulation platform designed by the International Water Association (IWA), enabling researchers to conduct experiments on a unified platform [31]. BSM1 simulates the modern urban wastewater treatment process, mainly the removal of nitrogen-containing organic matter. Figure 3 shows the basic structure of BSM1, which consists of a five-compartment activated sludge reactor (two anoxic tanks followed by three aerobic tanks) and a secondary settling tank. Different operation conditions are simulated by changing the influent data of the platform, such as during sunny days, rainy days, and rainstorm days.

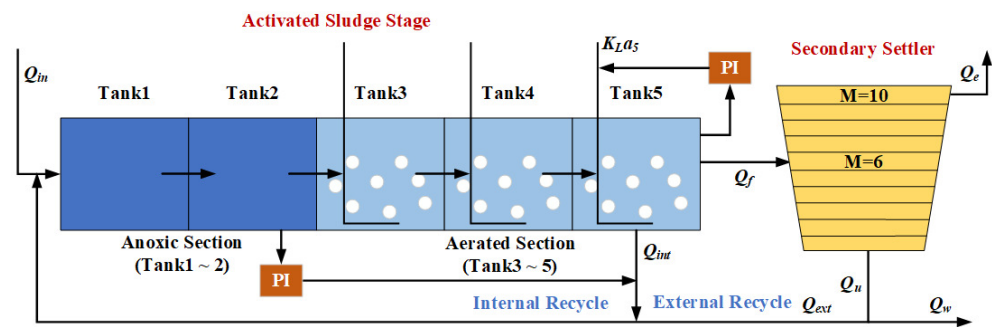


Figure 3. The basic structure of the BSM1 plant.

In this case, to verify the stability of the proposed soft sensor prediction performance, we collected experimental data under different weather conditions into a complex dataset containing multiple operation conditions. This dataset comprised data gathered from five sunny weekdays, a sunny weekend, two rainy days, and three rainstorm days. The process variables were collected at different stages including water inlet, reaction tanks (tank 1 to tank 5), secondary settler, recycle, and water outflow. An introduction to the specific variables is provided in Table 1. The sampling period for each variable was 15 min, which clearly reflects the changes in pollutants and treatment efficiency. We used the data from the first three sunny days (288 samples) as initial training data, and the remaining nine days of data (864 samples) as test data to assess the efficacy of the models.

Table 1. The notation and definition of process variables and system components.

Notation	Definition	Notation	Definition
SI	Soluble inert organic matter	XI	Particulate inert organic matter
SS	Readily biodegradable substrate	XS	Slowly biodegradable substrate
SO	Oxygen	XBH	Active heterotrophic biomass
SNO	Nitrate and nitrite nitrogen	XBA	Active autotrophic biomass
SNH	$NH_4^+ + NH_3$ nitrogen	XP	Particulate products arising from biomass decay
SND	Soluble biodegradable organic nitrogen	XND	Particulate biodegradable organic nitrogen
SALK	Alkalinity	TSS	Total readily biodegradable substrate
Q	Flow rate	BOD	Biochemical oxygen demand
in	Water inlet	bef	Summary input before tank 1
r1	Reaction tank 1	int	Internal recycle
r2	Reaction tank 2	feed	Input of the secondary settler
r3	Reaction tank 3	ext	External recycle
r4	Reaction tank 4	under	Underflow
r5	Reaction tank 5	out	Water outflow

Figure 4 illustrates the curves of the process variables of different treatment stages in the WWTP, including input port, reaction tank 1, reaction tank 2, reaction tank 3, reaction tank 4, reaction tank 5, and secondary settler. We can see that corresponding fluctuations occur in the process variables in the subsequent treatment stages simultaneously due to variations in the influent data. For example, two obvious mutation points in the influent data are located at the 673rd and 870th sample points, and corresponding changes occur in the subsequent treatment stages at the same sample points. Furthermore, we find that preprocessing of the collected data can effectively eradicate the problem of time delay. In other words, these process variables sampled at different locations can reflect the wastewater treatment information at the same moment.

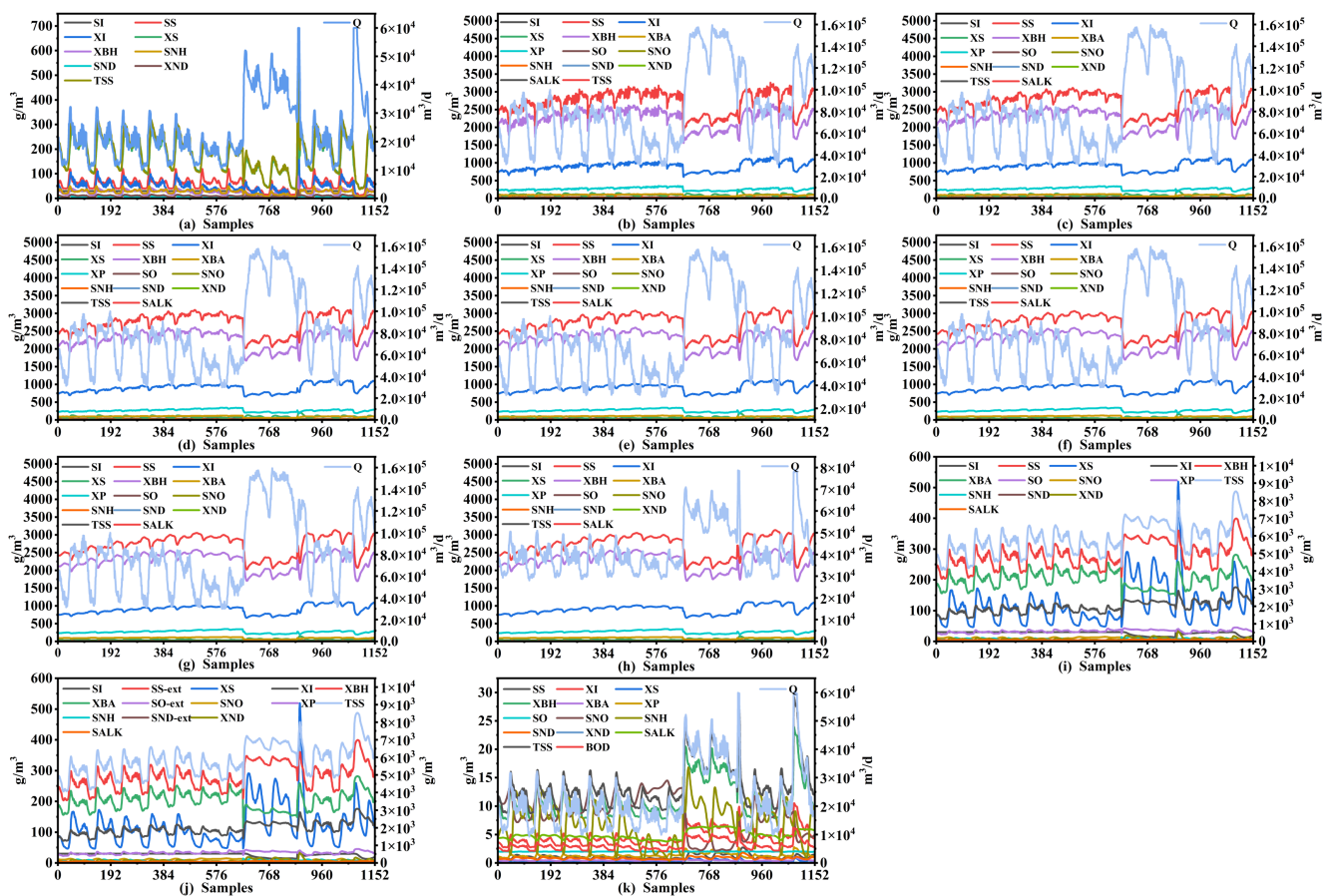


Figure 4. The curves of process variables in different stages of the BSM1 plant. (a) Original inflow of the plant; (b) input of reaction tank 1; (c) output of reaction tank 1; (d) output of reaction tank 2; (e) output of reaction tank 3; (f) output of reaction tank 4; (g) output of reaction tank 5; (h) input of the secondary settler; (i) underflow of the secondary settler; (j) external recycle; (k) outflow of the plant.

3.2. Variable Selection Results

In this case, we collected a total of 157 process variables, which cover various stages of the global wastewater treatment process. Utilizing all these variables for modeling would lead to a significant increase in model complexity and could adversely impact the prediction accuracy due to the accumulation of redundant information. Therefore, following the application of KNN-MI for variable selection, we selected 22 auxiliary variables from the 157 process variables. The selected auxiliary variables have strong correlations with effluent BOD, which can meet the accuracy requirements for predicting effluent BOD. At the same time, it reduces the complexity of the data dimensions and soft sensing models, eliminates information redundancy caused by weakly correlated variables, and is of great significance for improving the predictive performance of the model. Figure 5 shows the KNN-MI values between each process variable and effluent BOD, with a threshold line of $\lambda = 1.2$. When the KNN-MI value exceeds 1.2, it means that the variable has a strong correlation with the effluent BOD. On the contrary, when the KNN-MI value is below this threshold line, the variable is removed.

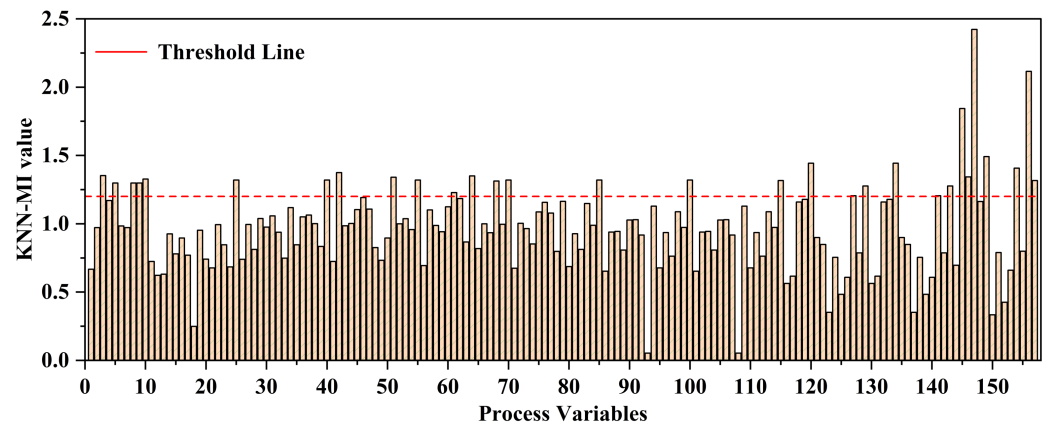


Figure 5. The KNN-MI values between each process variable and effluent BOD.

3.3. Evaluation Indicators

To verify the predictive performance of the proposed soft sensor, root mean square error (RMSE), mean absolute error (MAE), and Pearson correlation coefficient (PCC) are selected as evaluation metrics [32]. The specific formulas are as follows:

$$RMSE = \frac{1}{N} \sqrt{\sum_{i=1}^N (\hat{y}_i - y_i)^2} \quad (15)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i| \quad (16)$$

$$PCC(Y, \hat{Y}) = \frac{Cov(Y, \hat{Y})}{\sqrt{Var(Y) \cdot Var(\hat{Y})}} \quad (17)$$

where N represents the number of test data. y_i and $\hat{y}_i$ are the real value and predictive value of the i -th sample. Y and $\hat{Y}$ are the real and predictive datasets. $Cov(\cdot)$ is the covariance and $Var(\cdot)$ is the variance.

3.4. Prediction Results and Analysis

The global performance evaluation indicators of the BPNN, CNN, Bi-LSTM, MW-Bi-LSTM, and the proposed MPDAR-Bi-LSTM models are shown in Table 2. Firstly, for all the five models, it can be seen that the prediction accuracy of the three models based on Bi-LSTM is far superior to BPNN and CNN, while the time consumption is longer than CNN and shorter than BPNN. Secondly, for Bi-LSTM, MW-Bi-LSTM, and the proposed MPDAR-Bi-LSTM models, the traditional Bi-LSTM model exhibits inferior predictive performance compared to the other two models. This disparity is primarily ascribed to the inherent characteristics of Bi-LSTM as a prevalent deep neural network, typically employed for offline training. In the face of frequent variations in conditions prompted by an intricate external environment, offline models often demonstrate limited adaptability, particularly in instances of significant fluctuations such as inclement weather conditions. In contrast, the MW-Bi-LSTM and the proposed MPDAR-Bi-LSTM models demonstrate notably enhanced predictive capabilities compared to the Bi-LSTM model. This can be attributed to their capacity for sustaining robust model performance through continual model reconstruction. Nonetheless, it is essential to acknowledge that the reconstruction process leads to increased time consumption. As illustrated in Table 2, the time consumption of the two models is 1068.31 s and 638.47 s, respectively, significantly exceeding the 34.65 s recorded for the Bi-LSTM model. However, for the two prediction methods with the ability to update the

model, the MPDAR-Bi-LSTM model has a selective update mechanism due to the MPDAR strategy, unlike the MW-Bi-LSTM model. For example, as shown in Figure 6a, the 1–96 test samples on sunny days have similar distribution characteristics with the initial training data; the change rate of $E(X)$ is J , calculated using the MPDAR strategy, and J does not surpass the designated threshold, so there is no need to update the MPDAR-Bi-LSTM prediction model. But the MW-Bi-LSTM model lacks an analysis of the data distribution characteristics and continuously updates the prediction model. So, the proposed MPDAR-Bi-LSTM model is clearly more efficient. Therefore, in summary, the proposed MPDAR-Bi-LSTM model has higher model accuracy and stronger robustness compared to the traditional Bi-LSTM model, BPNN model, and CNN model, and has better overall performance than the MW-Bi-LSTM model, while consuming less time.

Table 2. The global performance evaluation indicators of the five different models.

Model	RMSE (g/m ³)	MAE (g/m ³)	PCC	Time (s)
BPNN	0.5769	0.3260	0.1649	12.63
CNN	0.9973	0.7137	0.6704	35.39
Bi-LSTM	0.4023	0.2269	0.5556	34.65
MW-Bi-LSTM	0.1464	0.0831	0.9736	1068.31
MPDAR-Bi-LSTM	0.1392	0.0752	0.9762	638.47

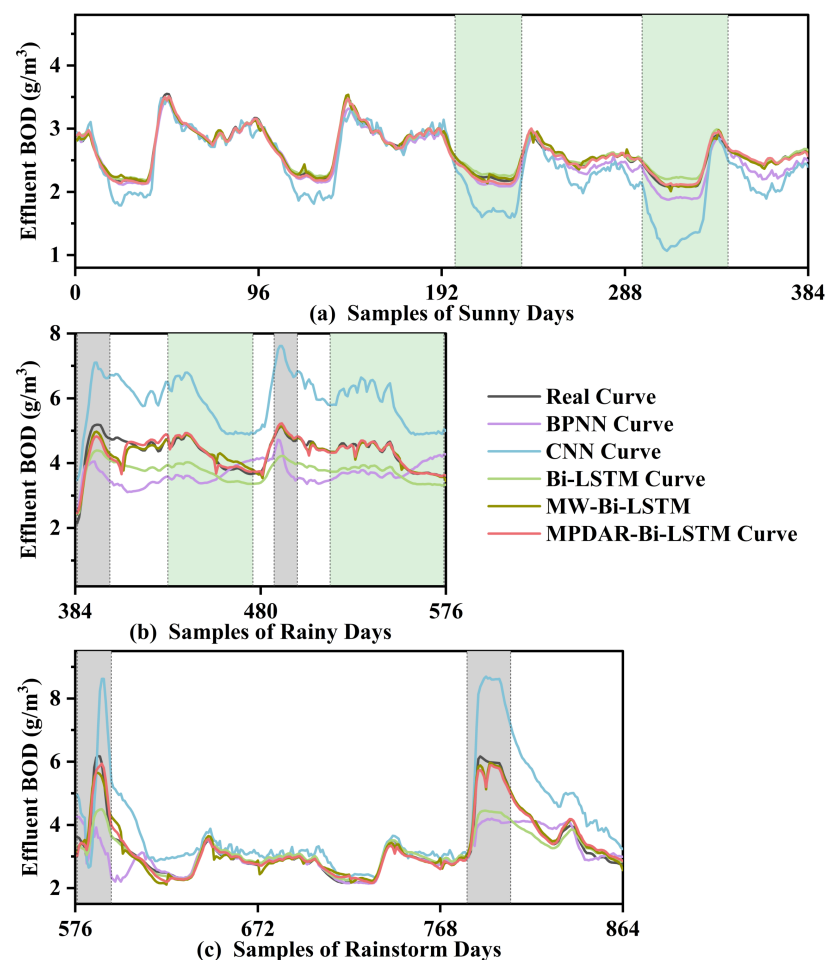


Figure 6. Prediction results of effluent BOD in the WWTP under (a) sunny days, (b) rainy days, and (c) rainstorm days. (For interpretation of the colors in this figure, the reader is referred to the web version of this article.)

In order to compare the prediction results of the above-mentioned five different models more intuitively, Figure 6 shows the fitting result between the predictive curve and the real curve of effluent BOD in WWTP under different weather conditions; Table 3 shows the performance evaluation indicators of the five models under different weather conditions. As shown in Figure 6a, the first two days are sunny weekdays, and the process data are close to the initial training data. The prediction performance of BPNN, Bi-LSTM, MW-Bi-LSTM, and the proposed MPDAR-Bi-LSTM models is good, but the prediction error of the CNN model is a little large. The last two days are sunny weekend days, and there are slight changes in the working conditions compared to sunny weekdays. The prediction performance of the BPNN and CNN models deteriorates. Due to the inability to update the prediction model, the prediction performance of the Bi-LSTM model is slightly worse than the MW-Bi-LSTM and the MPDAR-Bi-LSTM models, as shown in the green shaded area. According to the performance indices in Table 3, it can be seen that the MPDAR-Bi-LSTM model has better overall performance compared to the MW-Bi-LSTM model. As shown in Figure 6b, there were significant changes in the working conditions on rainy days compared to sunny weekdays. The BPNN, CNN, and Bi-LSTM models had a large prediction error due to their inability to update the model, while the MW-Bi-LSTM model and the MPDAR-Bi-LSTM model were able to update the prediction model in a timely manner, achieving better prediction results. However, upon further analysis, the MW-Bi-LSTM model could make corresponding adjustments more quickly after switching operating conditions compared to the MPDAR-Bi-LSTM model, as shown in the gray shaded area of Figure 6b. Although the adjustment time of the MPDAR-Bi-LSTM model is slightly longer, the final prediction result is better than the MW-Bi-LSTM model, as shown in the green shaded area in Figure 6b. Finally, as shown in Figure 6c, the working conditions on rainstorm days changed more than those on sunny weekdays. The BPNN, CNN, and Bi-LSTM models cannot update the model, leading to significant prediction errors. The MW-Bi-LSTM model and the MPDAR-Bi-LSTM model can update the prediction model in time to achieve better prediction results. Different from the previous two situations, the change in working conditions on rainstorm days is more extreme. In the early stage, the fast tracking performance of the MPDAR-Bi-LSTM model is slightly weaker than that of MW-Bi-LSTM, as shown in the gray shaded area of Figure 6c. But as time passes, the adjusted MPDAR-Bi-LSTM model gradually achieved more accurate prediction results. In addition, under the gentle working conditions between the two extreme rainstorm events, the tracking performance of the MPDAR-Bi-LSTM model is better than that of MW-Bi-LSTM. Therefore, the proposed MPDAR-Bi-LSTM model has the best comprehensive performance under normal working conditions, unstable working conditions, and extreme events.

Table 3. The performance evaluation indicators of the five different models under different conditions.

Method	Sunny Days			Rainy Days			Rainstorm Days		
	RMSE ¹	MAE ²	PCC ³	RMSE	MAE	PCC	RMSE	MAE	PCC
BPNN	0.1049	0.0815	0.9185	0.9157	0.8215	−6.5879	0.6517	0.3216	−0.2092
CNN	0.3456	0.2547	0.6066	1.6136	1.5819	−3.3784	1.0434	0.7468	0.5724
Bi-LSTM	0.0384	0.0323	0.9995	0.7602	0.5928	0.9652	0.4756	0.2423	0.9673
MW-Bi-LSTM	0.0356	0.0300	0.9996	0.2458	0.1198	0.9973	0.1875	0.1295	0.9957
MPD-Bi-LSTM	0.0307	0.0284	0.9997	0.2623	0.1212	0.9969	0.1608	0.1068	0.9968

¹ The units of RMSE are g/m³. ² The units of MAE are g/m³. ³ PCC is a dimensionless coefficient.

The experimental results show that the proposed MPDAR-Bi-LSTM soft sensor exhibits the capability to optimize the prediction precision and efficiency in the face of complex wastewater treatment processes. This achievement is primarily attributed to the MPDAR strategy. This strategy calculates the statistical mean values of various procedural variables and integrates them according to the KNN-MI values, thereby offering a more thorough portrayal of operational conditions of wastewater treatment plants. Then, by leveraging

the variation rate of the mathematical expectation, it determines whether the operational conditions have changed, and selectively reconstructs the model. However, it is not difficult to find that even though the proposed MPDAR strategy largely avoids ineffective updates, it still requires a significant amount of time cost contrasted with offline models. Moreover, when faced with extreme changes, new models perform slightly poorly in predicting extreme values during the initial prediction phases. Finally, the "memory function" of Bi-LSTM neural networks may sometimes transmit negative information, thereby influencing the accuracy of prediction performance.

4. Conclusions

In order to improve the prediction accuracy of effluent BOD in the wastewater treatment process and ensure the timeliness of prediction results, this paper proposed an MPDAR-Bi-LSTM soft sensor. The proposed soft sensor firstly uses KNN-MI to select appropriate process variables for modeling, thereby reducing data dimensionality. Then, a Bi-LSTM neural network with "memory function" is used to construct the prediction model. Afterwards, to enhance model robustness and prevent ineffective model reconstruction, a novel MPDAR strategy is developed to update the model. Finally, experiments are conducted based on the dataset, with complex operational conditions, collected from BSM1. The experimental results show that compared with BPNN and CNN methods, the overall predictive performance of methods based on Bi-LSTM is better because each process variable in the wastewater treatment processes is affected by its own temporal changes and the prediction models based on Bi-LSTM can better extract the latent information during data transmission. As for the three methods based on Bi-LSTM, compared to the traditional Bi-LSTM method, the proposed MPDAR-Bi-LSTM model significantly improves the prediction accuracy. And compared with the MW-Bi-LSTM method, although the adjustment time of the proposed model is slightly longer, the total time consumption is reduced by 40.24% and the RMSE and MAE of the final prediction result is improved by 4.92% and 9.51%, respectively. To sum up, the proposed MPDAR-Bi-LSTM model has higher model accuracy and stronger robustness compared to the traditional BPNN, CNN, and Bi-LSTM models, and has better overall performance than the MW-Bi-LSTM model, while consuming less time. But then again, it cannot be denied that there is still a lot of room for improvement in the proposed method, and future research directions may include exploring large-scale models and diverse adaptive algorithms.

Author Contributions: Conceptualization, W.L. and Y.L.; methodology, W.L. and D.L.; software, D.L.; validation, D.L. and J.Z.; formal analysis, W.L. and D.L.; investigation, D.L. and J.Z.; resources, Y.L.; data curation, W.L. and D.L.; writing—original draft preparation, W.L. and D.L.; writing—review and editing, W.L., Y.L., D.L. and J.Z.; visualization, W.L. and D.L.; supervision, Y.L.; project administration, Y.L.; funding acquisition, Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China grant number 62273362.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

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

References

1. Francesco, D.C.; Simona, D.S.; Ferraro, A.; Petrella, A. Phosphorous removal and recovery from urban wastewater: Current practices and new directions. *Sci. Total. Environ.* **2022**, *823*, 153750–153768.
2. Santos, A.F.; Alvarenga, P.; Gando-Ferreira, L.M.; Quina, M.J. Urban wastewater as a source of reclaimed water for irrigation: Barriers and future possibilities. *Environments* **2023**, *10*, 17. [[CrossRef](#)]

3. Li, D.; Yang, C.; Li, Y.; Zhou, C.; Huang, D.; Liu, Y. A deep semi-supervised learning framework towards multi-output soft sensors development and applications in wastewater treatment processes. *J. Water Process. Eng.* **2024**, *57*, 104654–104665. [\[CrossRef\]](#)
4. Zhou, M.; Zhang, Y.; Wang, J.; Shi, W.T.; Puig, V. Water quality indicator interval prediction in wastewater treatment process based on the improved BES-LSSVM algorithm. *Sensors* **2022**, *22*, 422. [\[CrossRef\]](#)
5. Li, W.; Han, J.; Li, Y.; Zhang, F.; Zhou, X.; Yang, C. Optimal sensor placement method for wastewater treatment plants based on discrete multi-objective state transition algorithm. *J. Environ. Manag.* **2022**, *307*, 114491–114502. [\[CrossRef\]](#)
6. Liu, Y.; Sun, Y.; Fang, G.; Ye, H.; Cai, Z. Targeting wastewater quality variables prediction: Improving sparrow search algorithm towards optimizing echo state network. *J. Water Process. Eng.* **2024**, *65*, 105717–105731. [\[CrossRef\]](#)
7. Perera, Y.S.; Ratnaweera, D.; Dasanayaka, C.H. The role of artificial intelligence-driven soft sensors in advanced sustainable process industries: A critical review. *Eng. Appl. Artif. Intell.* **2023**, *121*, 105988–106012. [\[CrossRef\]](#)
8. Nair, A.; Hykkerud, A.; Ratnaweera, H. Estimating phosphorus and COD concentrations using a hybrid soft sensor: A case study in a Norwegian municipal wastewater treatment plant. *Water* **2022**, *14*, 332. [\[CrossRef\]](#)
9. Andrade, P.; Silva, I.; Silva, M.; Flores, T.; Cassiano, J.; Costa, D.G. A tinyml soft-sensor approach for low-cost detection and monitoring of vehicular emissions. *Sensors* **2022**, *22*, 3838. [\[CrossRef\]](#)
10. Meng, X.; Zhang, Y.; Qiao, J. An adaptive task-oriented RBF network for key water quality parameters prediction in wastewater treatment process. *Neural Comput. Appl.* **2021**, *33*, 11401–11414. [\[CrossRef\]](#)
11. Pattnaik, B.S.; Pattanayak, A.S.; Udgate, S.K.; Panda, A.K. Machine learning based soft sensor model for BOD estimation using intelligence at edge. *Complex Intell. Syst.* **2021**, *7*, 961–976. [\[CrossRef\]](#)
12. Liu, W.; Lin, S.; Li, X.; Li, W.; Deng, H.; Fang, H.; Li, W. Analysis of dissolved oxygen influencing factors and concentration prediction using input variable selection technique: A hybrid machine learning approach. *J. Environ. Manag.* **2024**, *357*, 120777–120788. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Xiao, B.; Li, Y.; Sun, B.; Yang, C.; Huang, K.; Zhu, H. Decentralized PCA modeling based on relevance and redundancy variable selection and its application to large-scale dynamic process monitoring. *Process Saf. Environ. Prot.* **2024**, *151*, 85–100. [\[CrossRef\]](#)
14. Mokhtari, S.; Abbaspour, A.; Yen, K.K.; Sargolzaei, A. A machine learning approach for anomaly detection in industrial control systems based on measurement data. *Electronics* **2021**, *10*, 407. [\[CrossRef\]](#)
15. Li, D.; Yang, C.; Li, Y.; Chen, Y.; Huang, D.; Liu, Y. Learning a neural network-based soft sensor with double-errors parallel optimization towards effluent variable prediction in wastewater treatment plants. *J. Environ. Manag.* **2024**, *366*, 121907–121916. [\[CrossRef\]](#)
16. Ou, C.; Zhu, H.; Shardt, Y.A.W.; Ye, L.; Yuan, X.; Wang, Y.; Yang, C. Quality-driven regularization for deep learning networks and its application to industrial soft sensors. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, 1–11. [\[CrossRef\]](#)
17. Wang, S.; Ren, P.; Takyi-Aninakwa, P.; Jin, S.; Fernandez, C. A critical review of improved deep convolutional neural network for multi-timescale state prediction of lithium-ion batteries. *Energies* **2022**, *15*, 5053. [\[CrossRef\]](#)
18. Qiao, J.; Lin, Y.; Bi, J.; Yuan, H.; Wang, G.; Zhou, M. Attention-Based Spatiotemporal Graph Fusion Convolution Networks for Water Quality Prediction. *IEEE Trans. Autom. Sci. Eng.* **2024**, 1–10. [\[CrossRef\]](#)
19. Suo, Y.; Chen, W.; Claramunt, C.; Yang, S. A ship trajectory prediction framework based on a recurrent neural network. *Sensors* **2020**, *20*, 5133. [\[CrossRef\]](#)
20. Abdar, M.; Pourpanah, F.; Hussain, S.; Rezazadegan, D.; Liu, L.; Ghavamzadeh, M. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Inf. Fusion* **2021**, *76*, 243–297. [\[CrossRef\]](#)
21. Xu, W.; Liu, P.; Cheng, L.; Zhou, Y.; Xia, Q.; Gong, Y.; Liu, Y. Multi-step wind speed prediction by combining a WRF simulation and an error correction strategy. *Renew. Energy* **2021**, *163*, 772–782. [\[CrossRef\]](#)
22. Zhu, X.; Rehman, K.U.; Wang, B.; Shahzad, M. Modern soft-sensing modeling methods for fermentation processes. *Sensors* **2020**, *20*, 1771. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Guo, F.; Wei, B.; Huang, B. A just-in-time modeling approach for multimode soft sensor based on Gaussian mixture variational autoencoder. *Comput. Chem. Eng.* **2021**, *146*, 107230–107238. [\[CrossRef\]](#)
24. Liu, B.; Cai, Y.; Guo, Y.; Chen, X. Transtailor: Pruning the pre-trained model for improved transfer learning. *Proc. AAAI Conf. Artif. Intell.* **2021**, *35*, 8627–8634. [\[CrossRef\]](#)
25. Li, D.; Yang, C.; Li, Y. A multi-subsystem collaborative Bi-LSTM-based adaptive soft sensor for global prediction of ammonia-nitrogen concentration in wastewater treatment processes. *Water Res.* **2024**, *254*, 121347–121361. [\[CrossRef\]](#)
26. Chen, T.L.; Fushing, H.; Chou, E.P. Learned Practical Guidelines for Evaluating Conditional Entropy and Mutual Information in Discovering Major Factors of Response-vs.-Covariate Dynamics. *Entropy* **2022**, *24*, 1382. [\[CrossRef\]](#)
27. Wang, W.; Yang, C.; Han, J.; Li, W.; Li, Y. A soft sensor modeling method with dynamic time-delay estimation and its application in wastewater treatment plant. *Biochem. Eng. J.* **2021**, *172*, 108048–108057. [\[CrossRef\]](#)
28. Recio-Colmenares, R.; León Becerril, E.; Gurubel Tun, K.J.; Conchas, R.F. Design of a Soft Sensor Based on Long Short-Term Memory Artificial Neural Network (LSTM) for Wastewater Treatment Plants. *Sensors* **2023**, *23*, 9236. [\[CrossRef\]](#)
29. Alfwzan, W.F.; Selim, M.M.; Althobaiti, S.; Hussin, A.M. Application of Bi-LSTM method for groundwater quality assessment through water quality indices. *J. Water Process. Eng.* **2023**, *53*, 103889–103902. [\[CrossRef\]](#)
30. Wang, C.; Li, Y.; Huang, K.; Yang, C.; Gui, W. VAE4RSS: A VAE-based neural network approach for robust soft sensor with application to zinc roasting process. *Eng. Appl. Artif. Intell.* **2022**, *114*, 105180–105197. [\[CrossRef\]](#)

31. Han, H.; Wang, Y.; Liu, Z.; Sun, H.; Qiao, J. Knowledge-Data Driven Optimal Control for Nonlinear Systems and Its Application to Wastewater Treatment Process. *IEEE Trans. Cybern.* **2024**, *54*, 6132–6144. [[CrossRef](#)] [[PubMed](#)]
32. Li, D.; Liu, Y.; Huang, D.; Lui, C.F.; Xie, M. Development of an adversarial transfer learning based soft sensor in industrial systems. *IEEE Trans. Instrum. Meas.* **2023**, *79*, 1–10. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.