Data-Driven Management for Fuzzy Sewage Treatment Processes Using Hybrid Neural Computing

Wenru Zeng [1], Zhiwei Guo [1], Yu Shen [1], Ali Kashif Bashir [2], Keping Yu [3*], Yasser D. Al-Otaibi [4], Xu Gao [1,5*]

[1] National Research Base of Intelligent Manufacturing Service, Chongqing Technology and Business University, Chongqing 400067, China
[2] Department of Computing and Mathematics, Manchester Metropolitan University, UK
[3] Global Information and Telecommunication Institute, Waseda University Tokyo, Japan
[4] Department of Information Systems in Rabigh, King Abdulaziz University, Jeddah 21589, Saudi Arabia
[5] Chongqing Sino French Environmental Excellence Research & Development Center Co. Ltd., Chongqing 400042, China

Correspondence: keping.yu@aoni.waseda.jp (Keping Yu) and 2018065@ctbu.edu.cn (Xu Gao)

Abstract. With the growing public attention on sustainable development and green ecosystems, the efficient management of fuzzy sewage treatment processes (FSTP) has been a major concern in academia. Characterized by strong abstraction and analysis abilities, data mining technologies provide a novel perspective to solve this problem. In recent years, data-driven management for FSTP has been widely investigated, resulting in a number of typical approaches. However, almost all existing technical approaches consider FSTP a unidirectional, sequential process, ignoring the bidirectional temporality caused by backflow operations. Therefore, we propose a data-driven management mechanism for FSTP based on hybrid neural computing (IM-HNC for short). This mechanism attempts to capture the bidirectional time-series features of FSTP with the aid of a bidirectional long short-term memory model, and further introduces a convolutional neural network to construct feature spaces with a stronger expression capability. Empirically, we implement a series of experiments on three datasets under different parameter settings to test the efficiency and robustness of the proposed IM-HNC. The experimental results manifest that the IM-HNC has an average performance improvement of approximately 5% compared to the baselines.

Keywords: data-driven management, fuzzy sewage treatment process (FSTP), hybrid neural computing, bidirectional time-series features, green ecosystems

1 Introduction

As sustainable development is held in high regard today, green ecosystems will become an important prospect for future smart cities [1, 2]. This gives rise to the exploration of new techniques or management patterns in related fields, the most typical of which is the fuzzy sewage treatment process (FSTP) [3–6]. The FSTP is actually a kind of typical and complex industrial scenarios, in which treatment process is implemented accompanied with much fuzziness and uncertainty [7–9]. Owing to its importance to achieving green ecosystem, the realization of its optimal management is highly correlated to the sustainable engineering of cities [10–12]. The core of management for FSTP is to control the amount of dissolved oxygen (DO) in intermediate processes so that the index values of pollutants can be reduced [13–15]. Because it is difficult to directly determine an appropriate amount of DO,

precise advance predictions for sewage treatment results will conversely contribute to decision making [16–18]. But the conventional fuzzy analysis method is not suitable to deal with this kind of problem [19–21]. To this end, modeling schemes that can well express the entire FSTP from input to output is of great significance [22–24].

However, there are great difficulties in achieving the efficient management for FSTP, and accurately predicting the results of FSTP is challenging [25, 26]. From the perspective of systems science, the complex biochemical processes inside FSTP are characterized by nonlinearity, uncertainty and hysteresis [27, 28]. In particular, chemicals must undergo unperceivable and continuous reactions during treatment processes, which makes efficient management dramatically challenging [29–31]. Currently, the popularity of the Internet of Things in FSTP facilitates the collection of a large amount of monitoring data during these industrial processes [32, 33]. The massive amount of data can be used to achieve more refined management for FSTP through high-performance computation [34–36]. In particular, in imagined future cities, intelligent-algorithm-embedded management mechanisms for FSTP are particularly important [37–39].

At present, many researchers have attached importance to data-driven management for FSTP. In general, the biochemical process of sewage treatment is regarded as a gray box system without considering the biochemical mechanisms inside [40, 41]. For instance, Pang et al. [42] proposed a management pattern to implement multi-water quality monitoring based on a support vector machine. Cong et al. [43] suggested an adaptive learning method based on a wavelet neural network to optimize the management of FSTP. Some researchers have also managed to build models for FSTPs by considering time-series features [44, 45]. For example, Li et al. [46] proposed an intelligent management pattern based on the sparse auto-encoder and the long short-term memory model (LSTM). However, existing management patterns regarded FSTP as a unidirectional sequential process [47, 48], ignoring the fact that time series features are actually bidirectional due to the existence of backflow in FSTP. Specifically, sewage flows in both the forward and backward directions during long-term treatment processes.

To solve the aforementioned challenge, this paper uses the bidirectional long short-term memory (Bi-LSTM) model to capture the bidirectional time-series features of FSTP and introduces a convolutional neural network (CNN) model to enhance feature expression. Therefore, we propose a data-driven Intelligent Management mechanism based on Hybrid Neural Computing (IM-HNC) for FSTP. The core of this approach is to predict the treatment results of FSTP by formulating a data mining-based model. In particular, a CNN is developed to

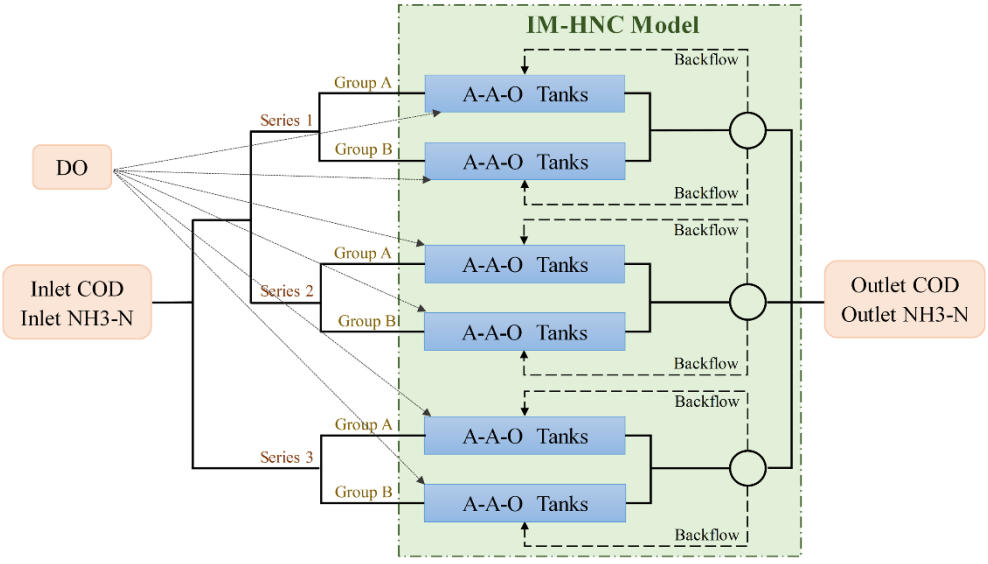


Fig. 1 Process structure of FSTP studied in this paper

construct feature spaces for the initial features of an FSTP, and the Bi-LSTM is leveraged to express the long-term temporality of the FSTP. The proposed IM-HNC deeply represents both the forward and backward sequential features of the FSTP and thus more appropriately expresses the entire FSTP through hybrid neural computing. In addition, a series of experiments are carried out to test the performance of the proposed IM-HNC on three datasets split from a real dataset.

The remainder of this paper is organized as follows: The questions and framework of this research are illustrated in Section 2. The detailed mathematical processes of the method are described in Section 3. In Section 4, data preprocessing and extensive experiments are implemented to assess the performance and stability for IM-HNC, and conclusions are presented in Section 5.

2 General Outline

2.1 Description of Problem

The experimental dataset in this paper was collected via biochemical sensors from a sewage treatment plant using an A^2/O process in Chongqing, China. The A^2/O process is an industrial sewage solution technique that has evolved from the original activated sludge process [49]. And A^2 refers to Anaerobic-Anoxic, O is Oxic [50]. Fig. 1 presents the process structure of FSTP studied in this paper, and the core terms are defined as follows:

Definition 1 (Inlet COD): The chemical oxygen demand (COD) amount before entering FSTP

Definition 2 (Inlet NH_3-N): The ammonia nitrogen (NH_3-N) amount before entering FSTP.

Definition 3 (DO): The DO amount added during the A^2/O process.

Definition 4 (Outlet COD): The COD amount after completion of FSTP.

Definition 5 (Outlet NH_3-N): The NH_3-N amount after completion of FSTP.

A long-term sewage treatment process can be viewed as a collection of operations at separate timestamps, and COD and NH_3-N are selected as the major pollutants monitored in this case. The inlet conditions and outlet results with respect to these two pollutants are denoted as x_1, x_2 and y_1, y_2, respectively. It can be observed from Fig. 1 that there are six main oxic tanks in the treatment process. The DO monitoring values inside the tanks are denoted as x_3 to x_8. Given the inlet conditions and the intermediate DO amount, the main goal of this paper is to develop a mapping model to calculate the outlet results in advance. Table 1 displays the main indicators of FSTP studied in this paper.

Table 1 The main indicators of FSTP studied in this paper

Variable	Definition
x_1, x_2	Inlet COD and Inlet NH_3-N
x_3, x_4	DO in oxic tanks of Series 1
x_5, x_6	DO in oxic tanks of Series 2
x_7, x_8	DO in oxic tanks of Series 3
y_1, y_2	Outlet COD and Outlet NH_3-N

2.2 Framework

Fig. 2 describes the overall architecture of the IM-HNC pattern, which contains three main modules: the convolution module, the sequential characteristics module and the full connection module. The inlet conditions and DO at the t-th timestamp are expressed as $x_i(t)$ $(i = 1,2,\cdots,8;\ t = 1,2,\cdots,T)$, where $i = 1,2$ denote the inlet COD and the inlet NH_3-N, respectively, and $i = 3,4,\ldots,8$ denote the DO density values. The total $x_i(t)$ and $y_j(t)$ are subjected to correlation analysis, and the parts with strong correlations are selected. The retained $x_i(t)$ can be finally aggregated into a sequence feature $X(t)$ that will be inputted into the convolution module to be encoded into $Con(t)$. Then, the $Con(t)$ is regarded as the input of each timestamp in the sequential characteristics module and further encoded into $H(t)$. Subsequently, the output $H_L(t)$ of the sequential characteristics module undergoes dimension reduction and decoding through the full connection module. Finally, the real-world dataset is inputted the IM-HNC to make it predict sewage treatment results, and make recommendations for FSTP management.

Therefore, given the inlet parameter indexes and the DO configuration at the same timestamp, the outlet parameter indexes will be generated according to the IM-HNC.

3 Methodology

3.1 Convolutional Neural Network Modeling

The convolutional neural network is mainly responsible for feature extraction due to its strong ability of abstract information representation. As illustrated in Fig. 3, its architecture mainly contains two groups: a convolutional layer and a max-pooling layer. The former maps feature vectors of previous stages into higher-order feature representations via convolution operations, and the latter appropriately reduces the dimensionality of the feature representations by pooling operations [51].

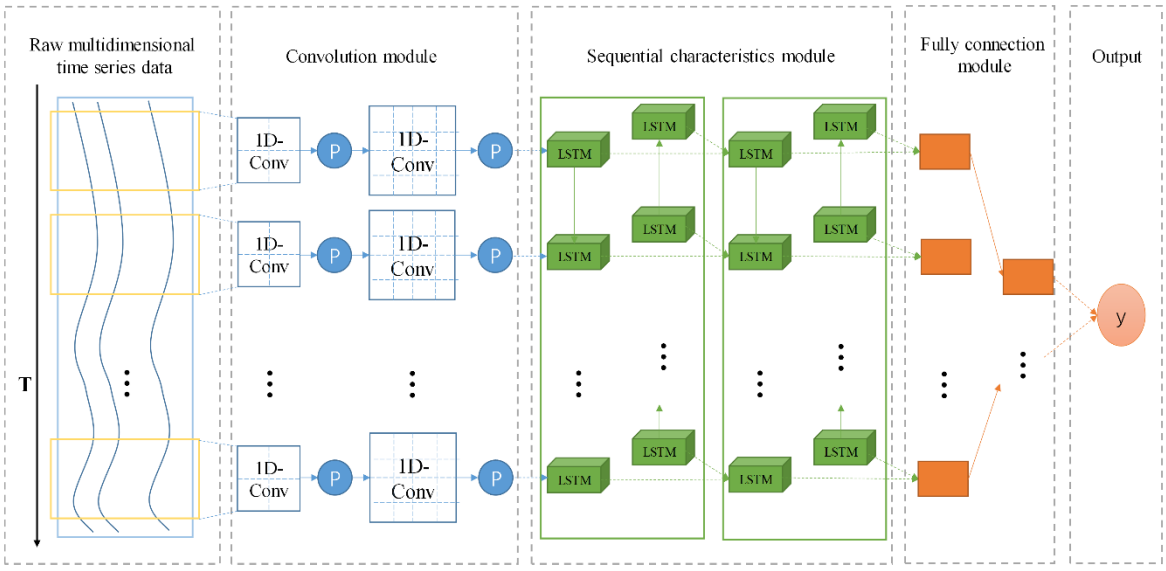


Fig. 2 An overview of the architecture of the IM-HNC pattern

Convolution refers to an inner product operation between feature vectors or matrices and filtering matrices; the first convolutional layer has 24 filters, and the second convolutional layer has 48 filters. This can be expressed in the following formula:

$$h_i(t) = relu(h_{i-1}(t) \otimes \omega_i(t) + b_1(t)) \tag{1}$$

where $h_i(t)$ is the feature vector of the i-th convolution layer at the t-th timestamp, $\otimes$ is the convolution operation of the convolution kernel, $\omega_i(t)$ refers to weight vector of the i-th layer convolution kernel, and $relu(\cdot)$ is the ReLU activation function, which is calculated as

$$relu(x) = max(0, x) \tag{2}$$

The role of pooling is to select a rule to compress the input feature map, thereby reducing the dimension of the output values [52]. The pooling rule used in this article is max-pooling, which selects the maximum value output to the next level for the values within the $m \times m$-dimensional neighborhood in the feature map and is expressed as

$$p_i(t) = \max pooling\left(p_{i-1}(t)\right) \tag{3}$$

where $p_i(t)$ is the feature vector of the i-th pooling layer at the t-th timestamp, and $maxpooling(\cdot)$ is the pooling rule. The kernel of the two max-pooling layers in the convolution module is set to 2×2 dimensions. Finally, the output $Con(t)$ of the convolution module at the t-th timestamp is inputted into the sequential characteristics module for encoding.

3.2 Bidirectional Long Short-Term Memory Modeling

The bidirectional long short-term memory model is an advanced form of the long short-term memory model. As represented in Fig. 4, the operation of the Bi-LSTM processes both forward and backward LSTMs at the same time. In addition, the LSTM updates the network status via three gating units, which are illustrated as follows:

- *Forget gate (FG):* The forget gate determines the extent to which the current moment $C(t)$ retains the previous moment state of the unit $C(t-1)$. When the forget gate is in the open state, the data in the storage unit can be saved; when the gate is in the closed state, the data in the unit is cleared.

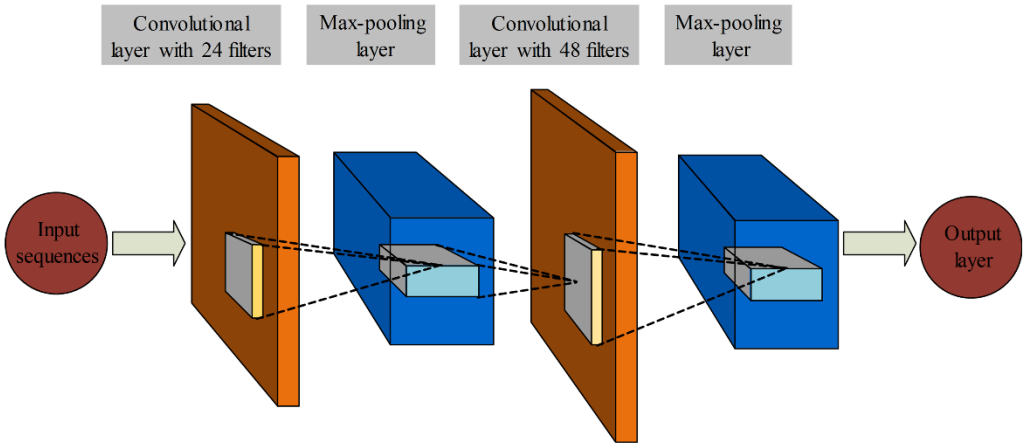


Fig. 3 Structure diagram of the convolution module

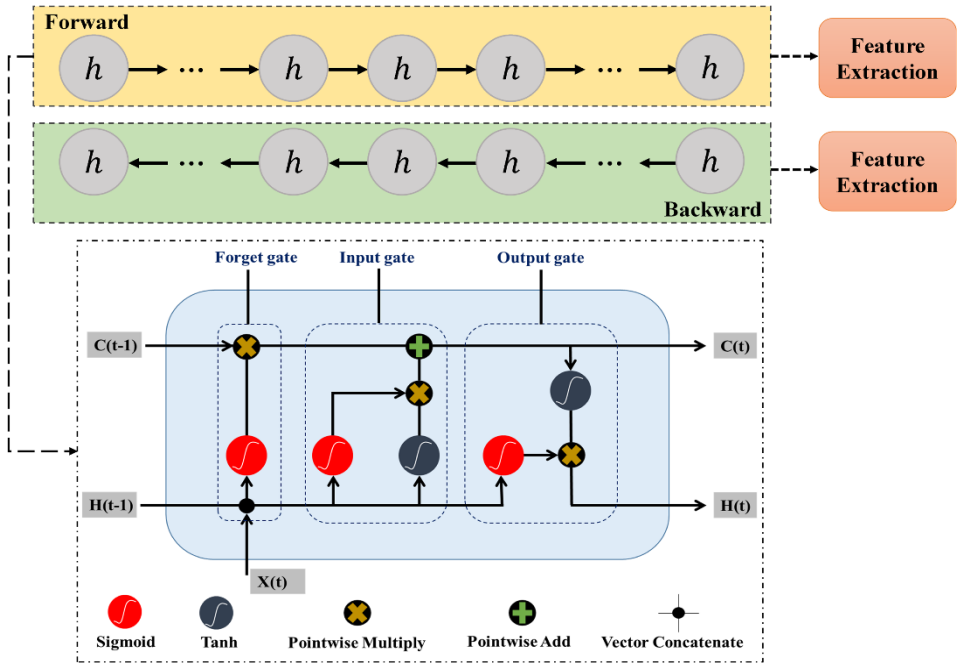


Fig. 4 The structure of the sequential characteristics module

- *Input gate (IG):* The input gate determines the amount of network input at the current time $X(t)$ saved in the unit state $C(t)$. As the input gate is opened, the storage unit can identify the contents of other neurons as part of the memory; when the input gate is closed, the storage unit cannot recognize these contents as memory.
- *Output gate (OG):* The output gate determines the value of the desired output $\mathrm{H(t)}$ via the unit state $\mathrm{C(t)}$. When the output gate is on, other neurons can discern the contents stored in the storage unit; when it is off, the other neurons cannot discern these contents.

The LSTM takes the received signal $X(t)$ and output $H(t-1)$ at the t-1-th timestamp as inputs and obtains the forget factor $f(t)$ after FG, which is expressed as

$$f(t) = sigmoid(W_f[H(t-1), X(t)] + b_f(\mathrm{t})) \tag{4}$$

where W_f is the connection weight matrix between the input and the FG; $b_f(t)$ is the connection bias at the t-th timestamp; and $sigmoid(\cdot)$ refers to the $sigmoid$ activation function, which can be expressed as the following formula:

$$sigmoid = \frac{1}{1 + exp\ (-x)} \tag{5}$$

When $f(t)$ is equal to 0, the historical information in the unit is completely forgotten; when the value of $f(t)$ is 1, the historical information is remembered. Therefore, the IG updates the status according to the input. The unit history information is calculated as follows:

$$i(t) = \delta(W_i[H(t-1), X(t)] + b_i(\mathrm{t})) \tag{6}$$

$$\tilde{C}(t) = tanh\ (W_c[H(t-1), X(t)] + b_c(\mathrm{t})) \tag{7}$$

$$C(t) = f(t) \cdot C(t-1) + i_t \cdot \tilde{C}(t) \tag{8}$$

Among Equations (6-8), W_i and W_c are the IG connection weight and the $tanh$ layer weight, respectively; b_i and b_c refer to the corresponding offsets; and $\tilde{C}(t)$ is the candidate value generated by the $tanh$ layer that

participates in the update process of unit information $C(t)$. The output control factor $O(t)$ is obtained from the OG, as expressed in Equation (9), which decides whether or not to unite the output information. The output $H(t)$ at the t-th timestamp is illustrated in Equation (10):

$$O(t) = \delta(W_o[H(t-1), X(t)] + b_o(t)) \tag{9}$$

$$H(t) = O(t) \cdot tanh\,(C(t)) \tag{10}$$

Based on the increased accuracy of the combined prediction determined by the previous input and the future input, a bidirectional recurrent neural network is proposed. Although the Bi-LSTM requires multiple trainings with input information to obtain convergence, it has a high accuracy and the same gate unit as the LSTM. To understand the Bi-LSTM, imagine an input sequence L with n elements. The order of the forward LSTM is expressed as $\{L_1, L_2, \dots, L_{n-1}, L_n\}$, while that of the backward LSTM is the opposite. The forward and backward LSTMs separately undergo the training process first. The training results of these LSTMs are integrated through fusion, which can be expressed as

$$H(t) = H_F(t) \circledast H_B(n-t+1) \tag{11}$$

where H_F and H_B refer to the outputs of the forward and backward LSTMs, respectively, and $\circledast$ is any integration operator, such as a simple accumulation. The outputs of the sequential characteristics module are fed in the fully connection module to predict sewage treatment results.

The proposed IM-HNC in this paper is an intelligent management pattern based on a hybrid neural computing of the CNN and the Bi-LSTM. The configuration details of this pattern are shown in Table 2.

Table 2 Parameter setting details of the IM-HNC

#NO	Layer Type	Output Shape	Param
1	Convolution1D	(None,6,24)	192
2	MaxPooling1D	(None,3,24)	0
3	Convolution1D	(None,3,48)	1200
4	MaxPooling1D	(None,1,48)	0
5	Bi-LSTM	(None,1,128)	57856
6	Dropout	(None,1,128)	0
7	Bi-LSTM	(None,64)	41216
8	Dropout	(None,64)	0
9	Fully connected layer	(None,32)	2080
10	Fully connected layer	(None,1)	33

3.3 Decoding and Optimization

The output vector $H(t)$ generated by each group in the sequential characteristics module is summarized as

$$H_l(t) = \{H_1(t), H_2(t), \dots, H_L(t)\} \tag{12}$$

Then, the full connection module is exploited as a part of the decoding to predict sewage treatment results.

$$\phi_l(t) = relu\{\delta[H_l(t)]\} \tag{13}$$

$$\varsigma_l(t) = relu\{\gamma[\phi_l(t)]\} \tag{14}$$

where $\delta(\cdot)$ and $\gamma(\cdot)$ are two full connection layers to output a single value. The output is a linear transformation of $\varsigma_l(t)$, which is calculated as

$$\hat{y}_l(t) = sigmoid\{W_l(t) \cdot \varsigma_l(t) + b_l(t)\} \tag{15}$$

where $W_l(t)$ and $b_l(t)$ are the weight vector and the bias vector at the t-th timestamp, respectively, and $\hat{y}_l(t)$ refers to the output of from the full connection module. To avoid overfitting, the loss function is the following optimization objective:

$$Q = \frac{1}{2}\sum_{t=1}^{T}(\hat{y}_\theta(t) - y(t))^2 \tag{16}$$

$$\theta_j \leftarrow \theta_j - \alpha\frac{\partial}{\partial\theta_j}J(\theta) \tag{17}$$

where $\hat{y}_\theta(t)$ and $y(t)$ denote the prediction values and the observed values of sewage treatment results, respectively, θ denotes a conventional parameter in gradient descent method, α is the learning rate, and $\frac{\partial}{\partial\theta_j}J(\theta)$ is the partial derivative of the loss function (Q) to the parameter θ_j. The purpose of the optimization is to screen out the parameters set that minimizes Q. The optimization method exploited here is the stochastic gradient descent (SGD) [53]. SGD randomly selects a group from the sample, updates it by gradient after training, and then selects a group and updates it again. Details of the iterative process are omitted due to the limited space of the paper.

After that, a complete intelligent management pattern is established for the outlet parameter indexes in the FSTP. Therefore, once the inlet parameter indexes and the DO configuration are inputted, the predicted results of the outlet parameter indexes at the same timestamp will be acquired correspondingly.

Table 3 Statistical analysis and summary of sewage treatment data

Variables	Min	Max	Mean	S. D	C. V
x_1 (mg/L)	9.339	1061.544	441.008	195.165	0.443
x_2 (mg/L)	0.156	110.467	27.283	9.889	0.362
x_3 (mg/L)	1.002	9.562	2.810	1.519	0.540
x_4 (mg/L)	1.000	9.287	3.293	1.861	0.565
x_5 (mg/L)	1.001	9.413	2.607	1.117	0.429
x_6 (mg/L)	1.000	9.088	2.691	1.109	0.412
x_7 (mg/L)	1.003	9.956	5.571	2.604	0.467
x_8 (mg/L)	1.241	9.973	6.152	3.118	0.507

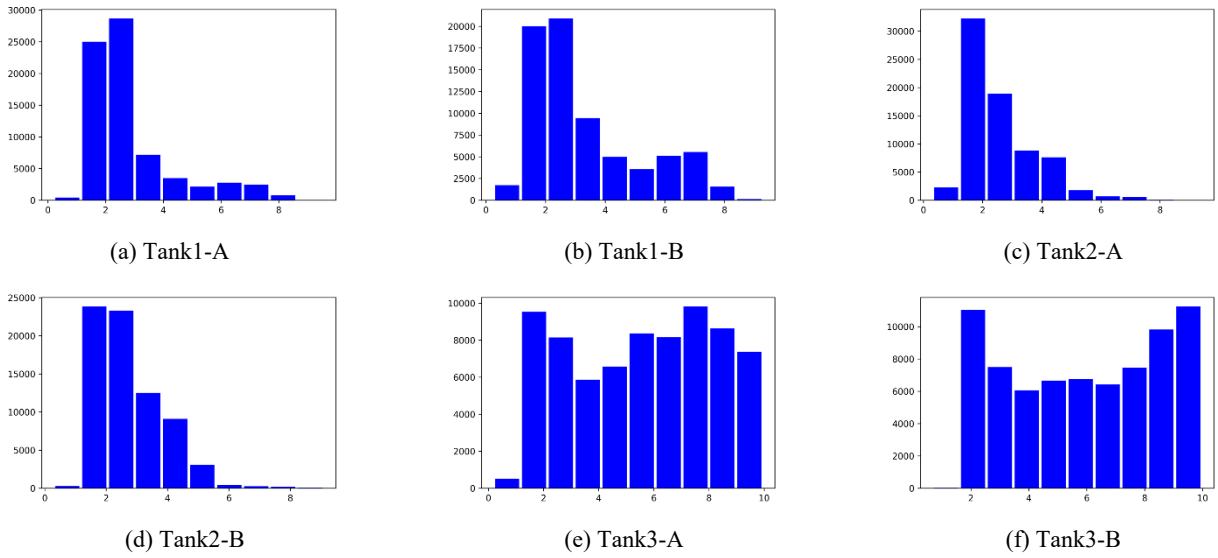
4 Experiments

4.1 Data Pre-processing

During the operation of a sewage treatment plant, data anomalies are often caused by technical problems that cannot be ruled out in time or by monitoring equipment failure. Therefore, it is necessary to preprocess the original data of the sewage treatment plant. The specific steps states are as follows:

- *Data review:* Analyze whether or not the distribution of the data falls on a reasonable interval and whether or not the data values are consistent with the actual situation.

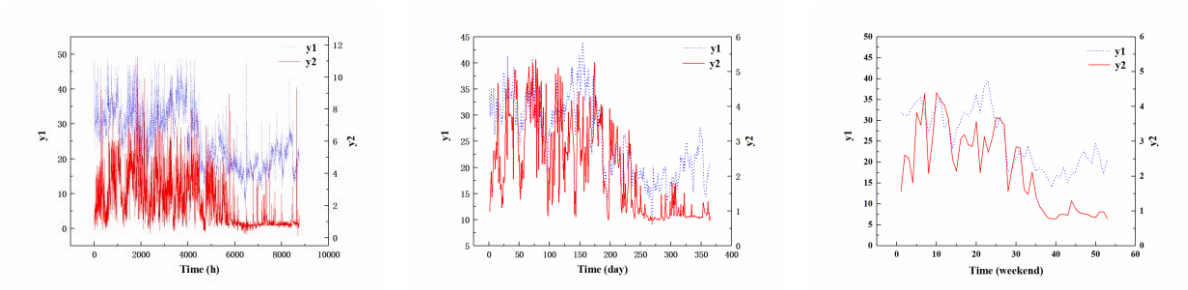
- *Missing statistics and processing:* For data missing due to equipment failure is executed by adding random number. The range of the random number is set to the maximum value of the sensor's monitoring data for the same day.
- *Data summarization:* The processed data are summarized, and the statistical analysis and summary are illustrated in Table 3. Fig. 5 shows the DO density distribution of the six oxic tanks in the A^2/O craft. The six subgraphs in the graphic represent Tank1-A, Tank1-B, Tank2-A, Tank2-B, Tank3-A and Tank3-B, respectively.



(a) Tank1-A (b) Tank1-B (c) Tank2-A

(d) Tank2-B (e) Tank3-A (f) Tank3-B

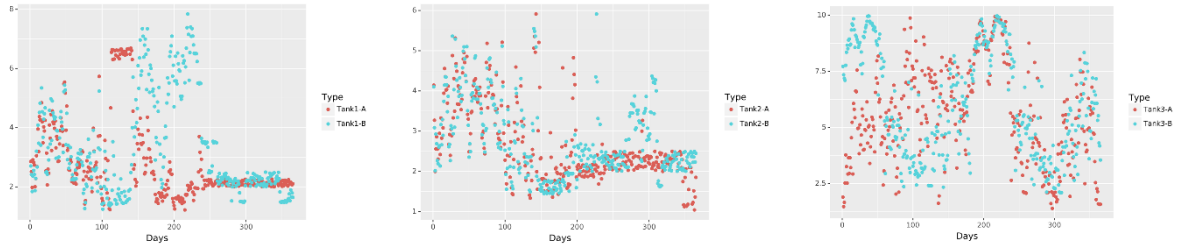
Fig. 5 The DO density distribution of the six oxic tanks in the A^2/O process



(a) Hourly dataset (b) Daily dataset (c) Weekly dataset

Fig. 6 The fluctuation of the outlet parameter indexes ($\boldsymbol{y_1}$, $\boldsymbol{y_2}$) in the three datasets

- *Data timing summary:* The aggregated datasets obtained in the third step are refined into hourly, daily, and weekly datasets. The fluctuations of the outlet parameter indexes (y_1, y_2) in the three datasets are shown in Fig. 6. Taking the daily dataset as an example, the sequence trend of the DO density values in group A and group B for each series of tanks can be observed from the three subgraphs in Fig. 7.
- *Correlation comparison:* The correlation analysis requires the correlation coefficient between each pair of variables to form a correlation coefficient matrix $R = [r_{ij}]_{p*q}$, where r_{ij} represents the correlation coefficient between the i-th variable x_i and the j-th variable y_j. A correlation coefficient greater than 0.1 was set as the threshold for selecting input variables. The results are shown in Table 4, and bold characters indicate compliance.



(a) The sequential trend of the DO in Tank 1

(b) The sequential trend of the DO in Tank2

(c) The sequential trend of the DO in Tank3

Fig. 7 The sequential trend of the DO density values

Table 4 The correlation coefficients (**R**) between individual dependent variables and $\mathbf{y_1}$, $\mathbf{y_2}$

Variables	R value between dependent variable and y_1	R value between dependent variable and y_2
x_1	**0.284**	**0.160**
x_2	0.099	**0.142**
x_3	**0.417**	**0.212**
x_4	**0.116**	**0.139**
x_5	**0.286**	**0.226**
x_6	**0.130**	0.098
x_7	0.058	0.087
x_8	**0.107**	**0.118**

4.2 Experimental Settings

In this phase, the generated prediction results in the previous step is evaluated. This article selects two performance metrics to assess the proposed intelligent management pattern IM-HNC: mean absolute error (MAE), root mean square error (RMSE). Their expressions are listed as follows:

$$MAE = \frac{1}{N}\sum_{o=1}^{N}|\hat{y}_o - y_o| \tag{18}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{o=1}^{N}(\hat{y}_o - y_o)^2} \tag{19}$$

where y_o and $\hat{y}_o$ denote the observed values and the prediction values of outlet indexes, respectively, and N is the total number of test set in data-driven pattern studied. For the above metrics, smaller values mean better model performance and accuracy. To compare the IM-HNC with several baseline methods from the perspective of stability and reliability. The selected several existing data-driven management methods are as follows:

1) Multi-layer Preceptor (MLP): It is a feedforward artificial neural networks in which every neuron is fully connected.
2) CNN: It is a kind of feedforward neural network with deep structure that includes convolution calculation.
3) LSTM: It is a kind of sequential neural network model specially designed to solve the long-term dependence problem.
4) CNN-LSTM: It is a hybrid neural network model combining CNN and LSTM in order to capture more complete features.

4.3 Results and Discussion

Two sets of experiments are carried out in this paper, one of which is to adjust the learning rate under the condition of given training set proportions and parameter values, and the other of which is to alter the training set proportion with fixed learning rates and parameter values.

In one group of experiments, the training set proportion and parameter value were set as default values, and the learning rate values were set to 0.01, 0.005 and 0.001.Table 5, Table 6 and Table 7 demonstrate the experimental results for the three datasets with different learning rate values. It can be observed in Table 5 that the IM-HNC predicts the outlet COD with a learning rate of 0.01, and the MAE and RMSE are 0.069 and 0.126, respectively; with a learning rate of 0.005, the MAE and RMSE are 0.068 and 0.106, respectively; with a learning rate of 0.001, the MAE and RMSE are 0.070 and 0.125, respectively. The prediction results of the outlet COD and outlet NH_3-N for the three datasets indicate that the values of the measurement indicators MAE and RSME were the lowest when the learning rate was 0.005. The data from the above table readily show that regardless of how the learning rate is adjusted, the IM-HNC proposed in this paper is superior to the other baselines methods in terms of both evaluation indicators.

Table 5 The performance of the four experimental methods for the hourly dataset with different learning rates

Dataset	Learning rate	Metrics	Outlet COD (y_1)					Outlet NH_3-N (y_2)				
			MLP	CNN	LSTM	CNN-LSTM	IM-HNC	MLP	CNN	LSTM	CNN-LSTM	IM-HNC
Hourly datasets	0.01	MAE	0.587	0.263	0.163	0.143	**0.069**	0.545	0.256	0.132	0.124	**0.074**
		RMSE	0.935	0.325	0.272	0.234	**0.126**	0.887	0.335	0.214	0.188	**0.148**
	0.005	MAE	0.455	0.243	0.156	0.141	**0.068**	0.404	0.251	0.125	0.123	**0.072**
		RMSE	0.778	0.305	0.237	0.210	**0.106**	0.748	0.329	0.193	0.164	**0.142**
	0.001	MAE	0.535	0.257	0.174	0.148	**0.070**	0.576	0.259	0.127	0.124	**0.074**
		RMSE	0.867	0.325	0.264	0.265	**0.125**	0.889	0.337	0.196	0.169	**0.144**

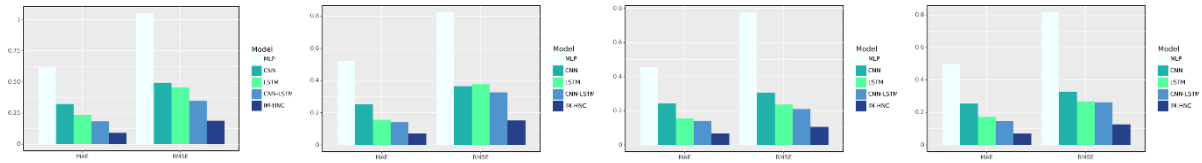
Table 6 The performance of the four experimental methods for the daily dataset with different learning rates

Dataset	Learning rate	Metrics	Outlet COD (y_1)					Outlet NH_3-N (y_2)				
			MLP	CNN	LSTM	CNN-LSTM	IM-HNC	MLP	CNN	LSTM	CNN-LSTM	IM-HNC
Daily datasets	0.01	MAE	0.558	0.283	0.182	0.153	**0.074**	0.563	0.264	0.137	0.124	**0.064**
		RMSE	0.749	0.375	0.325	0.255	**0.153**	0.650	0.346	0.234	0.195	**0.127**
	0.005	MAE	0.449	0.226	0.151	0.137	**0.061**	0.428	0.244	0.123	0.115	**0.056**
		RMSE	0.674	0.296	0.229	0.206	**0.095**	0.703	0.330	0.204	0.154	**0.092**
	0.001	MAE	0.533	0.254	0.165	0.149	**0.071**	0.525	0.247	0.128	0.169	**0.060**
		RMSE	0.729	0.314	0.286	0.225	**0.144**	0.755	0.368	0.215	0.185	**0.117**

Table 7 The performance of the four experimental methods for the weekly dataset with different learning rates

Dataset	Learning rate	Metrics	Outlet COD (y_1)					Outlet NH_3-N (y_2)				
			MLP	CNN	LSTM	CNN-LSTM	IM-HNC	MLP	CNN	LSTM	CNN-LSTM	IM-HNC
Weekly datasets	0.01	MAE	1.853	0.635	0.463	0.244	**0.163**	1.573	0.578	0.423	0.220	**0.149**
		RMSE	3.053	0.844	0.623	0.364	**0.278**	2.368	0.733	0.533	0.343	**0.235**
	0.005	MAE	1.683	0.536	0.357	0.189	**0.138**	1.274	0.426	0.329	0.153	**0.113**
		RMSE	2.415	0.744	0.533	0.275	**0.195**	2.147	0.634	0.473	0.238	**0.176**
	0.001	MAE	1.685	0.589	0.433	0.218	**0.159**	1.426	0.524	0.383	0.183	**0.139**
		RMSE	2.851	0.784	0.593	0.327	**0.239**	2.632	0.715	0.568	0.293	**0.203**

In another set of experiments, the learning rate and parameter value were set to default values, and the training set proportions were separately set to 30%, 50%, 70%, and 90%. Fig. 8, Fig. 10 and Fig. 12 illustrate the experimental results for predicting the outlet COD with different training set proportions for the three datasets. Fig. 9, Fig. 11 and Fig. 13 demonstrate the experimental results for predicting the outlet NH_3-N with different training set proportions for the three datasets. From Fig. 9, it can be clearly observed that the prediction results of the four models on the outlet NH_3-N always increase first and then gradually decrease with the increase of the training set proportion, and the proposed IM-HNC has the best prediction effect. The prediction results of the outlet COD and outlet NH_3-N for the three datasets show that the MAE and RMSE values for the four models were the lowest when the training set proportion was 70%, followed by 90%. The experimental results of this set were similar to those of the previous one: no matter how the training set proportions were adjusted, the IM-HNC had the best performances for both evaluation indicators.

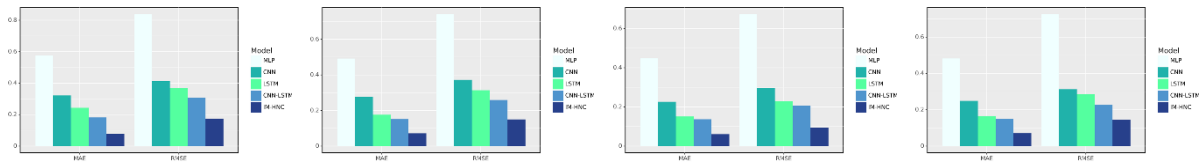


(a) training set:30% (b) training set:50% (c) training set:70% (d) training set:90%

Fig. 8 Comparison of the experimental results of predicting the outlet COD with different training set proportions for the hourly dataset

(a) Evaluation index: MAE (b) Evaluation index: RMSE

Fig. 9 Comparison of the experimental results of predicting the outlet NH_3-N with different training set proportions for the hourly dataset

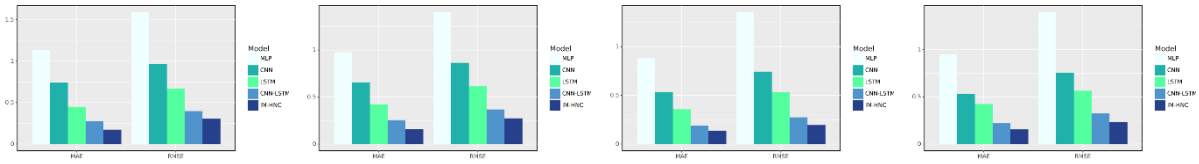


(a) training set:30% (b) training set:50% (c) training set:70% (d) training set:90%

Fig. 10 Comparison of the experimental results of predicting the outlet COD with different training set proportions for the daily dataset

(a) Evaluation index: MAE (b) Evaluation index: RMSE

Fig. 11 Comparison of the experimental results of predicting the outlet NH_3-N with different training set proportions for the daily dataset



(a) training set:30% (b) training set:50% (c) training set:70% (d) training set:90%

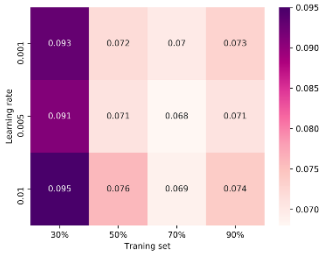
Fig. 12 Comparison of the experimental results of predicting the outlet COD with different training set proportions for the weekly dataset

(a) Evaluation index: MAE (b) Evaluation index: RMSE

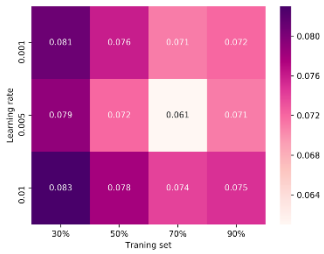
Fig. 13 Comparison of the experimental results of predicting the outlet NH_3-N with different training set proportions for the weekly dataset

A series of experimental results have confirmed the superiority of the performance of the proposed IM-HNC. The superior performance of the IM-HNC can be attributed to two factors. First, the CNN model was utilized to extract the underlying abstract features of the original data, and an effective feature space expression ability was acquired. Second, the Bi-LSTM model was adopted to accurately capture long-term bidirectional time-series features and generate a comprehensive representation of the global feature space.

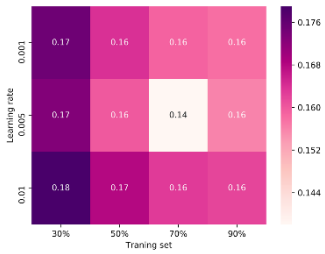
In addition, another special set of experiments was performed to assess the parameter sensitivity for IM-HNC. In this set of experiments, the IM-HNC was not compared with any foundation model; rather, its own results were compared after changing the learning rate and training set proportion. The MAE and RMSE results of the IM-HNC for predicting the outlet COD under different parameters are illustrated in Fig. 14 and Fig. 15, respectively. Fig. 16 and Fig. 17 show the MAE and RMSE results, respectively, of the IM-HNC for predicting the outlet NH_3-N under different parameter situations. Each of these four figures have three subgraphs, which correspond to the three types of datasets. It is obvious from the numerous subplots that the experimental results hardly changed after adjusting the parameters and datasets, which proves the stability of the proposed IM-HNC. The foremost reason for this phenomenon is that the IM-HNC integrates the effective characteristics of the FSTP, making it less susceptible to parameter changes.



(a) Changing of the learning rate and training set proportion for the hourly dataset

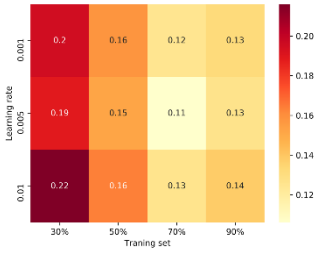


(b) Changing of the learning rate and training set proportion for the daily dataset

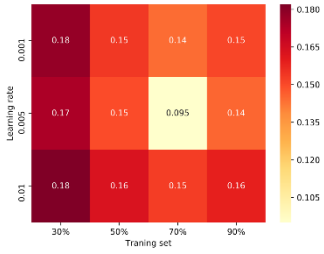


(c) Changing of the learning rate and training set proportion for the weekly dataset

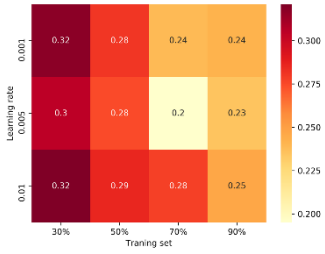
282 **Fig. 14** MAE results of the IM-HNC for predicting the outlet COD under different parameter situations



(d) Changing of the learning rate and training set proportion for the hourly dataset

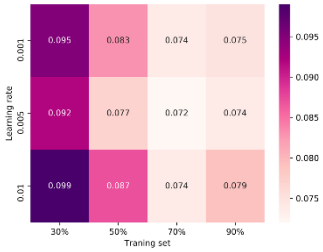


(e) Changing of the learning rate and training set proportion for the daily dataset

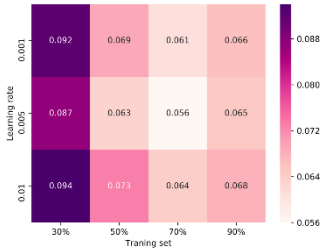


(f) Changing of the learning rate and training set proportion for the weekly dataset

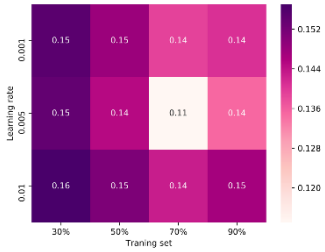
283 **Fig. 15** RMSE results of the IM-HNC with respect to the outlet COD under different parameter situations



(a) Changing of the learning rate and training set proportion for the hourly dataset

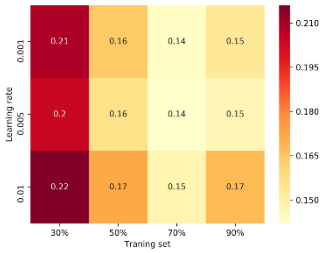


(b) Changing of the learning rate and training set proportion for the daily dataset

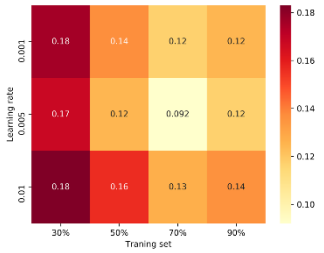


(c) Changing of the learning rate and training set proportion for the weekly dataset

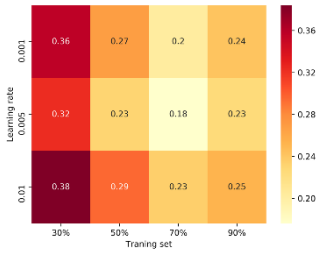
284 **Fig. 16** MAE results of the IM-HNC with respect to the outlet NH_3-N under different parameter situations



(a) Changing of the learning rate and training set proportion for the hourly dataset



(b) Changing of the learning rate and training set proportion for the daily dataset



(c) Changing of the learning rate and training set proportion for the weekly dataset

285 **Fig. 17** RMSE results of the IM-HNC with respect to the outlet NH_3-N under different parameter situations

In summary, the proposed IM-HNC in this paper had an excellent efficiency and stability for the forecast based on the hourly, daily, and weekly datasets and can effectively predict the relevant data of sewage treatment.

5 Conclusion

It is feasible to predict the results of sewage treatment via data analysis. However, the existing data-driven management pattern ignores the bidirectional time-series feature of FSTPs. Therefore, a new intelligent prediction and control management pattern, IM-HNC, is proposed in this paper to solve this problem. First, the convolutional module was exploited to automatically extract and encode the features of the FSTP. Next, the bidirectional time-series features of the FSTP are deeply expounded by the sequential characteristics module. Then, many experiments were implemented to verify the effectiveness and stability of the proposed IM-HNC on three datasets split from a real dataset. Finally, according to the experimental results, suggestions are provided for decision-making related to the intermediate process of the FSTP.

Acknowledgments.
This research was supported by National Key Research & Development Program of China (2016YFE0205600), Chongqing basic research and frontier exploration project of China (cstc2018jcyjAX0638), Chongqing Natural Science Foundation of China (cstc2019jcyj-msxmX0747), Scientific Program of Chongqing Technology and Business University (ZDPTTD201917, KFJJ2018071, 1952027), and Japan Society for the Promotion of Science (JSPS) Grants-in-Aid for Scientific Research (KAKENHI) under Grant JP18K18044.

References

1. Phan D, Bab-Hadiashar A, Lai CY, et al (2020) Intelligent energy management system for conventional autonomous vehicles. Energy 191:116476. https://doi.org/https://doi.org/10.1016/j.energy.2019.116476
2. Pawar P, TarunKumar M, Vittal K. P (2020) An IoT based Intelligent Smart Energy Management System with accurate forecasting and load strategy for renewable generation. Measurement 152:107187. https://doi.org/https://doi.org/10.1016/j.measurement.2019.107187
3. Zhou X, Li Y, Liang W (2020) CNN-RNN Based Intelligent Recommendation for Online Medical Pre-Diagnosis Support. IEEE/ACM Trans Comput Biol Bioinforma 1–1. https://doi.org/10.1109/tcbb.2020.2994780
4. Zhou X, Liang W, Wang KIK, et al (2018) Academic Influence Aware and Multidimensional Network Analysis for Research Collaboration Navigation Based on Scholarly Big Data. IEEE Trans Emerg Top Comput 1–12. https://doi.org/10.1109/TETC.2018.2860051
5. Zhou YM, Chen YP, Guo JS, et al (2019) Recycling of orange waste for single cell protein production and the synergistic and antagonistic effects on production quality. J Clean Prod 213:384–392. https://doi.org/10.1016/j.jclepro.2018.12.168
6. Zhou YM, Chen YP, Guo JS, et al (2019) The correlations and spatial characteristics of microbiome and silage quality by reusing of citrus waste in a family-scale bunker silo. J Clean Prod 226:407–418. https://doi.org/10.1016/j.jclepro.2019.04.075
7. Shafiq M, Tian Z, Bashir AK, et al (2020) Data mining and machine learning methods for sustainable smart cities traffic classification: A survey. Sustain Cities Soc 60:102177. https://doi.org/https://doi.org/10.1016/j.scs.2020.102177
8. Guo Z, Du B, Wang J, et al (2020) Data-driven prediction and control of wastewater treatment process through the combination of convolutional neural network and recurrent neural network. RSC Adv 10:13410–13419. https://doi.org/10.1039/d0ra00736f

9. Kramer FC, Shang R, Rietveld LC, Heijman SJG (2019) Fouling control in ceramic nanofiltration membranes during municipal sewage treatment. Sep Purif Technol 116373. https://doi.org/10.1016/J.SEPPUR.2019.116373
10. Tan L, Shi N, Yang C, Yu K (2020) A Blockchain-Based Access Control Framework for Cyber-Physical-Social System Big Data. IEEE Access 8:77215–77226. https://doi.org/10.1109/ACCESS.2020.2988951
11. Yang C, Tan L, Shi N, et al (2020) AuthPrivacyChain: A Blockchain-Based Access Control Framework with Privacy Protection in Cloud. IEEE Access 8:70604–70615. https://doi.org/10.1109/ACCESS.2020.2985762
12. Su J, Bai Q, Sindakis S, et al (2020) Vulnerability of multinational corporation knowledge network facing resource loss: A super-network perspective. Manag Decis. https://doi.org/10.1108/MD-02-2019-0227
13. Qin D, Yu J, Zou G, et al (2019) A Novel Combined Prediction Scheme Based on CNN and LSTM for Urban PM2.5 Concentration. IEEE Access 7:20050–20059. https://doi.org/10.1109/ACCESS.2019.2897028
14. Liu F, Zhou X, Wang T, et al (2019) An Attention-based Hybrid LSTM-CNN Model for Arrhythmias Classification. Proc Int Jt Conf Neural Networks 2019-July:1–8. https://doi.org/10.1109/IJCNN.2019.8852037
15. Fu Q, Niu D, Zang Z, et al (2019) Multi-stations' weather prediction based on hybrid model using 1D CNN and Bi-LSTM. Chinese Control Conf CCC 2019-July:3771–3775. https://doi.org/10.23919/ChiCC.2019.8866496
16. Jian J, Zhang Y, Jiang L, Su J (2020) Coordination of Supply Chains with Competing Manufacturers considering Fairness Concerns. Complexity 2020:. https://doi.org/10.1155/2020/4372603
17. Jian J, Guo Y, Jiang L, et al (2019) A multi-objective optimization model for green supply chain considering environmental benefits. Sustain 11:. https://doi.org/10.3390/su11215911
18. Yu K, Arifuzzaman M, Wen Z, et al (2015) A Key Management Scheme for Secure Communications of Information Centric Advanced Metering Infrastructure in Smart Grid. IEEE Trans Instrum Meas 64:2072–2085. https://doi.org/10.1109/TIM.2015.2444238
19. Alazab M, Alazab M, Shalaginov A, et al (2020) Intelligent mobile malware detection using permission requests and API calls. Futur Gener Comput Syst 107:509–521. https://doi.org/10.1016/j.future.2020.02.002
20. Tang M, Alazab M, Luo Y (2017) Big Data for Cybersecurity: Vulnerability Disclosure Trends and Dependencies. IEEE Trans Big Data 5:317–329. https://doi.org/10.1109/tbdata.2017.2723570
21. Etaher N, Weir GRS, Alazab M (2015) From ZeuS to zitmo: Trends in banking malware. Proc - 14th IEEE Int Conf Trust Secur Priv Comput Commun Trust 2015 1:1386–1391. https://doi.org/10.1109/Trustcom.2015.535
22. Zhou X, Hu Y, Liang W, et al (2020) Variational LSTM Enhanced Anomaly Detection for Industrial Big Data. IEEE Trans Ind Informatics 1–1. https://doi.org/10.1109/tii.2020.3022432
23. Zhou X, Liang W, Wang KIK, et al (2020) Deep-Learning-Enhanced Human Activity Recognition for Internet of Healthcare Things. IEEE Internet Things J 7:6429–6438. https://doi.org/10.1109/JIOT.2020.2985082
24. Su J, Wang J, Liu S, et al (2020) A Method for Efficient Task Assignment Based on the Satisfaction Degree of Knowledge. Complexity 2020:1–12. https://doi.org/10.1155/2020/3543782
25. Azab A, Alazab M, Aiash M (2016) Machine learning based botnet identification traffic. Proc - 15th IEEE Int Conf Trust Secur Priv Comput Commun 10th IEEE Int Conf Big Data Sci Eng 14th IEEE Int Symp Parallel Distrib Process with Appl IEEE Trust 2016 1788–1794. https://doi.org/10.1109/TrustCom.2016.0275
26. Azab A, Layton R, Alazab M, Oliver J (2015) Mining malware to detect variants. Proc - 5th Cybercrime Trust Comput Conf CTC 2014 44–53. https://doi.org/10.1109/CTC.2014.11
27. Li ZX, Renault FL, Gómez AOC, et al (2019) Nanofluids as secondary fluid in the refrigeration system: Experimental data, regression, ANFIS, and NN modeling. Int J Heat Mass Transf 144:118635. https://doi.org/10.1016/J.IJHEATMASSTRANSFER.2019.118635
28. Tang L, Lu X, Yang C, Li X (2019) Classification conducting knowledge acquisition by an evolutionary robust GRBF-NN model. Procedia Comput Sci 162:183–190. https://doi.org/10.1016/J.PROCS.2019.11.274
29. Song S, Wang L (2017) Modified GMDH-NN algorithm and its application for global sensitivity analysis. J Comput Phys 348:534–548. https://doi.org/10.1016/J.JCP.2017.07.027

30. Ceylan R, Koyuncu H (2019) A Novel Rotation Forest Modality Based on Hybrid NNs: RF (ScPSO-NN). J King Saud Univ - Comput Inf Sci 31:235–251. https://doi.org/10.1016/J.JKSUCI.2017.10.011
31. Thirumal Kumar D, Iyer S, Christy JP, et al (2019) A comparative computational approach toward pharmacological chaperones (NN-DNJ and ambroxol) on N370S and L444P mutations causing Gaucher's disease. Adv Protein Chem Struct Biol 114:315–339. https://doi.org/10.1016/BS.APCSB.2018.10.002
32. Guo Z, Shen Y AK et al. (2020) Robust Spammer Detection Using Collaborative Neural Network in Internet of Thing Applications. IEEE Internet Things J. https://doi.org/10.1109/JIOT.2020.3003802
33. Guo Z, Wang H (2020) A Deep Graph Neural Network-based Mechanism for Social Recommendations. IEEE Trans Ind Informatics 3203:1–1. https://doi.org/10.1109/tii.2020.2986316
34. Geng Z, Chen G, Han Y, et al (2020) Semantic relation extraction using sequential and tree-structured LSTM with attention. Inf Sci (Ny) 509:183–192. https://doi.org/10.1016/J.INS.2019.09.006
35. Wang K, Qi X, Liu H (2019) Photovoltaic power forecasting based LSTM-Convolutional Network. Energy 189:116225. https://doi.org/10.1016/J.ENERGY.2019.116225
36. Zhang B, Zhang H, Zhao G, Lian J (2020) Constructing a PM2.5 concentration prediction model by combining auto-encoder with Bi-LSTM neural networks. Environ Model Softw 124:104600. https://doi.org/10.1016/J.ENVSOFT.2019.104600
37. Liu W, Liu WD, Gu J (2020) Predictive model for water absorption in sublayers using a Joint Distribution Adaption based XGBoost transfer learning method. J Pet Sci Eng 188:106937. https://doi.org/10.1016/J.PETROL.2020.106937
38. Kazemi Y, Mirroshandel SA (2018) A novel method for predicting kidney stone type using ensemble learning. Artif Intell Med 84:117–126. https://doi.org/10.1016/j.artmed.2017.12.001
39. Zhang J, Yuan C, Wang C, et al (2020) Composite adaptive NN learning and control for discrete-time nonlinear uncertain systems in normal form. Neurocomputing. https://doi.org/10.1016/J.NEUCOM.2020.01.052
40. Abrougui K, Gabsi K, Mercatoris B, et al (2019) Prediction of organic potato yield using tillage systems and soil properties by artificial neural network (ANN) and multiple linear regressions (MLR). Soil Tillage Res 190:202–208. https://doi.org/10.1016/J.STILL.2019.01.011
41. Ghritlahre HK, Prasad RK (2018) Investigation of thermal performance of unidirectional flow porous bed solar air heater using MLP, GRNN, and RBF models of ANN technique. Therm Sci Eng Prog 6:226–235. https://doi.org/10.1016/J.TSEP.2018.04.006
42. Pang Z, Jia K (2013) Designing and accomplishing a multiple water quality monitoring system based on SVM. Proc - 2013 9th Int Conf Intell Inf Hiding Multimed Signal Process IIH-MSP 2013 121–124. https://doi.org/10.1109/IIH-MSP.2013.39
43. Cong Q, Yu W (2018) Integrated soft sensor with wavelet neural network and adaptive weighted fusion for water quality estimation in wastewater treatment process. Measurement 124:436–446. https://doi.org/10.1016/J.MEASUREMENT.2018.01.001
44. Jiang M, Zhang W, Zhang M, et al (2019) An LSTM-CNN attention approach for aspect-level sentiment classification. J Comput Methods Sci Eng 1:1–10. https://doi.org/10.3233/jcm-190022
45. Arshi S, Strachan R, Zhang L (2019) Weather Based Photovoltaic Energy Generation Prediction Using LSTM Networks. 2019 Int Jt Conf Neural Networks 14–19
46. Li Z, Peng F, Niu B, et al (2018) Water Quality Prediction Model Combining Sparse Auto-encoder and LSTM Network. IFAC-PapersOnLine 51:831–836. https://doi.org/10.1016/j.ifacol.2018.08.091
47. Hutt S (2017) Deep Learning for Cyber Security
48. Ebenuwa SH, Sharif MS, Alazab M, Al-Nemrat A (2019) Variance Ranking Attributes Selection Techniques for Binary Classification Problem in Imbalance Data. IEEE Access 7:24649–24666. https://doi.org/10.1109/ACCESS.2019.2899578
49. Zeng W, Guo Z WJ et al (2020) A Convolutional Neural Network-based Prediction Mechanism for Sewage Treatment. In: Proceedings - 2019 2nd International Conference on Communication, Network, and Artificial Intelligence, CNAI 2019. pp 1014–1021
50. Shaikh A (2018) Application of Microwaves in Sustainable Organic Synthesis. Green Chem 647–671. https://doi.org/10.1016/B978-0-12-809270-5.00023-6
51. Konstantinidis D, Argyriou V, Stathaki T, Grammalidis N (2020) A modular CNN-based building detector for remote sensing images. Comput Networks 168:107034. https://doi.org/10.1016/J.COMNET.2019.107034

52. Wang Z-R, Du J, Wang J-M (2020) Writer-aware CNN for parsimonious HMM-based offline handwritten Chinese text recognition. Pattern Recognit 100:107102. https://doi.org/10.1016/J.PATCOG.2019.107102
53. Nguyen LC, Nguyen-Xuan H (2020) Deep learning for computational structural optimization. ISA Trans. https://doi.org/https://doi.org/10.1016/j.isatra.2020.03.033