Neural Computing and Applications (2023) 35:7185–7206
https://doi.org/10.1007/s00521-021-06499-1

S.I. : NEURO, FUZZY AND THEIR HYBRIDIZATION

Knowledge and data-driven hybrid system for modeling fuzzy wastewater treatment process

Xuhong Cheng[1] · Zhiwei Guo[1] · Yu Shen[1,2] · Keping Yu[3] · Xu Gao[1,4]

Received: 2 December 2020 / Accepted: 30 August 2021 / Published online: 25 October 2021

Abstract
Since wastewater treatment processes (WTP) are generally accompanied with intense coupling and fuzziness, conventional biochemical mechanisms-based methods cannot comprehensively express the WTP due to limited computational ability. In response to the challenge caused by fuzziness, this paper proposes a hybrid control and prediction system for modeling WTP with the fuse of Activated Sludge model, Convolutional neural network and Long short-term memory neural networks (AS-CL) with knowledge and data-driven characteristics. Moreover, the activated sludge model is employed to model the wastewater treatment process based on the perspective of knowledge. Besides, the hybrid neural network that combines convolutional neural network and long short-term memory model is adopted to model the WTP from the perspective of data. Then, a multi-layer perception model is set up to realize collaborative awareness of data and knowledge. Lastly, the proposed AS-CL has been evaluated by a real-world data-set collected from a real sewage treatment plant. The results show that compared with typical existing methods, the proposal improves modeling efficiency. With the collaborative modeling scheme, influence from fuzziness on WTP can be reduced to some extent. Compared with five benchmark methods of the two evaluation indexes, the results of AS-CL show that the average performance of this method exceeds 7% of the baseline.

Keywords Hybrid system · Fuzzy wastewater treatment process · Data model · Knowledge model

1 Introduction

Despite the tremendous progress gained in the industrial era in the past few decades, the fuzziness of the industrial process still exists and wastewater treatment processes (WTP) are a typical type [1–3]. In recent years, it has become a hot topic to explore new technologies and management patterns in related fields. The model expression of WTP is able to guide the management of sewage treatment plants to a certain extent while improving treatment efficiency, especially the Internet of Things provides enough data support for modeling [4, 5]. The key indicator of WTP management is to control the amount of dissolved oxygen (DO) in reaction processes. The aeration device at the bottom of the aerobic tank supplies oxygen so that to reduce or remove the number of pollutants [6–8]. Therefore, the establishment of a model which can simulate the whole process of wastewater treatment from influent to effluent will extremely help to solve fuzziness in the process of WTP [9–11].

In recent years, efficient management mechanism of WTP has become a hot research point [12–15]. Currently, relevant management mechanisms can be divided into two categories: knowledge-driven methods and data-driven methods [16–20]. The former primary establishes the

✉ Yu Shen
shenyu@ctbu.edu.cn

✉ Keping Yu
keping.yu@aoni.waseda.jp

1 National Research Base of Intelligent Manufacturing Service, Chongqing Key Laboratory of Catalysis & New Environmental Materials, Chongqing Technology and Business University, Chongqing 400067, China

2 Chongqing South-to-Thais Environmental Protection Technology Research Institute Co., Ltd., Chongqing 400069, China

3 Global Information and Telecommunication Institute, Waseda University, Tokyo, Japan

4 Chongqing Sino French Environmental Excellence Research & Development Center Co., Ltd., Chongqing 400067, China

process model via researching the biochemical mechanism characteristics of WTP, such as Activated-Sludge-Models (AMS) series [21–23]. However, the knowledge-driven approaches are highly dependent on a great number of model parameters while the parameter measurement approaches are complicated [24, 25]. Because of the limitation of manual computation capability and increasing business volume at the present stage, results of knowledge-based methods are more and more difficult to satisfy the needs of industrial management. In contrast, data-driven methods aim to discover hidden features and patterns within the data from the perspective of statistics to enhance the level of automation and management efficiency [26, 27]. However, data-driven approaches to the industrial process of WTP have been highly abstracted, focusing on the statistical learning of data characteristics. Therefore, the impact of biochemical factors in WTP was completely ignored [28].

To overcome the limitations of knowledge-driven or data-driven methods, this paper attempts to capture efficient management solutions via knowledge and data-hybrid mechanism-driven method. In this paper, a unique hybrid system with the fuse of Activated Sludge model, Convolutional neural network and Long short-term memory neural networks (AS-CL) is proposed. This multi-mode integration [29, 30] will resolve the fuzziness associated with wastewater treatment processes more efficiently. As far as this paper can see, the scheme of combining the two mechanisms in the field of WTP to improve management in this field does not seem to be mentioned. In addition, plenty of existing research has focused on the modeling of processing processes with different time-stamps, which are mutually independent [31, 32]. However, a large number of dates have indicated that the changing process of the biochemical reaction will influence the treatment to process the next time-stamp. The proposed AS-CL in this paper is able to capture the global features of WTP more comprehensively, with better expression of the established process model. As far as this paper acknowledged, the proposed AS-CL is the first to implement knowledge-driven and data-hybridly driven management for WTP and consider the effect of multi-source feature fusion under a global perspective. Main contributions of this paper are summarized as follows:

(1) We reveal the limitations of single knowledge-driven methods and data-driven methods for WTP management.
(2) We put forward the AS-CL model with knowledge-driven and data-hybridly driven management mechanism for WTP under the internet of things environment.
(3) We carry out extensive experiments on three real-society data-sets to demonstrate the validity of the proposed AS-CL.

2 Related work

2.1 Data-driven modeling

The essence of data-driven modeling is to provide solutions for different industrial application requirements under different data characteristics and process characteristics. Many scholars studied and developed data-driven models to excavate the hidden features and patterns from the perspective of statistics to improve the level of industrial automation and management efficiency. For instance, Yaqub et al. [33] proposed and developed a neural network based on long short-term memory (LSTM) to predict nitrogen and phosphorus removal efficiency in an anaerobic-anoxic-oxic membrane bioreactor (A-A-O MBR) system. Wang et al. [34] proposed a dynamic chemical oxygen demand (COD) prediction model for urban sewage based on the hybrid CNN-LSTM deep learning algorithm to reduce unnecessary energy and material consumption. Nasser et al. [35] proposed a paradigm of the Internet of Things based on micro-services and containers to present a testbed for various scenarios that can be used in water resources management.

2.2 Knowledge-driven modeling

As a new research method, the activated sludge mathematical model has been applied widely in sewage treatment. Researchers at home and abroad have done considerable research on the activated sludge mathematical model, which mainly focuses on the utilization of activated sludge mathematical model to guide the design and transformation of the sewage treatment plant, optimization control, and so on. For example, in the study of Spérandio et al. [36], the submerged membrane bioreactor (SMBR) system model was established by using activated sludge model no.1 (ASM1) and activated sludge model no.3 (ASM3) models for the evaluation of municipal wastewater treatment. Kim et al. [37] adopted simplified activated sludge model no.2 (ASM2) to describe the dynamics of partial parameters in wastewater treatment plants. Chen et al. [24] established activated sludge model no.2d (ASM2d) to simulate the sewage treatment process, which was calibrated and validated for some parameters.

2.3 CNN-LSTM simulation modeling

The convolutional neural network (CNN) is a model widely used to construct data feature space, and long short-term memory (LSTM) models are mainly used to extract time-series features of data. The effective combination of the two can improve the accuracy and stability of the model to a certain extent. There are some typical examples. Wang et al. [34] proposed a prediction model for the dynamic chemical oxygen demand (COD) of urban sewage. The experimental results indicate that this hybrid CNN-LSTM deep learning algorithm has better prediction performance than the independent CNN or LSTM model. Li et al. [38] established a hybrid CNN-LSTM model to predict PM2.5 concentration in the next day. The experimental results show that the proposed multivariate CNN-LSTM has a prominent prediction effect compared with other models.

3 Overview

3.1 Problem statement

This research investigated a sewage treatment plant in Chongqing, which is a typical improved anaerobic-anoxic-oxic (A-A-O) process. It can be clearly observed from Fig. 1 that chemical oxygen demand (COD) and ammonia nitrogen (NH_4-N) are regarded as two primary pollution indexes in need of treatment. The biological treatment system has three series of improved A-A-O biological treatment processes in which each series contains two groups of anoxic-anaerobic-anoxic-oxic (A-A-A-O) treatment tanks with basically the same internal structure: a pre-anoxic tank, anaerobic tank, anoxic tank and oxic tank. Among them, underwater propeller is set in the anaerobic and anoxic tanks; membrane microporous equipment is installed in the oxic tank, and the air source provided via the blower room is used for aeration. Municipal sewage enters into the biological treatment system through a series of pre-treatment processes. The degradation process of the pollution index is achieved by adjusting the aeration rate, referring to the amount of dissolved oxygen (DO). In addition, the reflux condition of the A-A-A-O system is stipulated as follows: the process of the mixture from the oxic tank to the anoxic tank is regarded as internal reflux, and its internal reflux rate is 270%; the process of sludge from the secondary sedimentation tank to the pre-anoxic tank is taken as external reflux, and the external reflux rate is 100%. The experimental data-set employed in this research came from the major indexes in WTP monitored in a sewage treatment plant in 2018. The purpose of this paper is to establish an AS-CL mechanism to generate a mapping model, which predicts the export pollutant indexes according to the inlet pollutant indexes and the DO concentration value in each tank. Figure 1 demonstrates kernel process structure of the sewage treatment plant, and the core terms are defined as follows:

Definition 1 Chemical Oxygen Demand (COD): It refers to the amount of oxidant consumed to oxidize the reducing

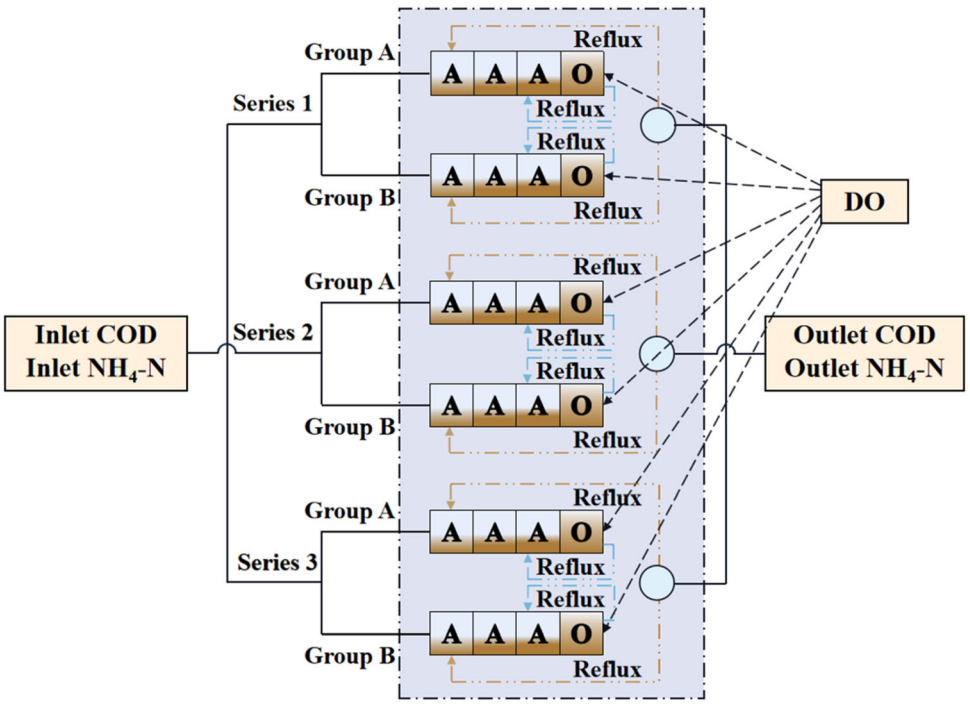


Fig. 1 Kernel process structure of the sewage treatment plant

substances in 1-L (liter) water sample under certain conditions, which can be converted into the number of milligrams of oxygen required for each liter of water sample to be fully oxidized. It reflects the degree to which the water is polluted for reducing substances.

Definition 2 Ammonia Nitrogen (NH_4-N): It refers to nitrogen in the form of free ammonia (FNH_3) and ammonium ions (NH_4^+) in water, that is, combined nitrogen in the form of ammonia or ammonium ions.

Definition 3 Dissolved Oxygen (DO): It refers to the oxygen in the air or aeration device that is dissolved in water to form dissolved oxygen, which is added to the aerobic tank.

3.2 Framework

Figure 2 illustrates framework of the proposed AS-CL mechanism, which contains two stages. The first stage contains two parts. In the first part, model simulation software is utilized to put up the ASM model for the actual process of the sewage treatment plant-based A-A-O theory. In the second part, CNN-LSTM hybrid neural network [39, 40] is introduced to construct a data model of wastewater treatment process from the perspective of statistics. In this part, the first learning is performed to train the CNN-LSTM model by inputting real-monitor data-sets to make it possess prediction ability. In the second stage, the prediction results of the ASM model and hybrid neural network model are integrated and re-predicted through a multi-layer perceptron [41, 42]. Therefore, the second learning is carried out in this stage, and the prediction result of the previous stage is used as the input data-set of this stage to train the CNN-LSTM model.

The inlet COD and the inlet NH_4-N at the t-th time-stamp are expressed as $\Re_{in}$ and $\mathfrak{B}_{in}$, respectively. The amount of DO added to oxic tanks at the t-th time-stamp is denoted as $\mathfrak{U}_{ij}^{(t)}$, where i denotes the six groups and j denotes the two oxic tanks of each group. The total parameters can be finally aggregated into a feature matrix $y_{(t)}$ which will be input into CNN model to be output into h_i. Then, the h_i is regarded as input of each time-stamp in LSTM [43] model and further output a feature vector F.

Besides, $y_{(t)}$ will be input into the ASM model to obtain the output result W_e. Finally, F and W_e are retrained by the multi-layer perceptron to further obtain the prediction result o_{t+1}^i.

In addition, the main symbols and acronyms used in this study are summarized in Tables 1 and 2

4 Methodology

This section illustrates the modeling process of the proposed AS-CL through three main parts. Firstly, the CNN was employed for extracting the data spatial feature; and LSTM was applied to extract the temporal features. Then, the activated sludge model is used to construct the microscopic biological model. Finally, a multi-layer perception model is established to realize collaborative awareness of data and knowledge. The description of the technology roadmap for this section is shown in Fig. 3.

4.1 Data modeling

In the process of wastewater prediction, water quality index exists two essential attributes: the spatial attribute which is the interrelations within wastewater input; and the time

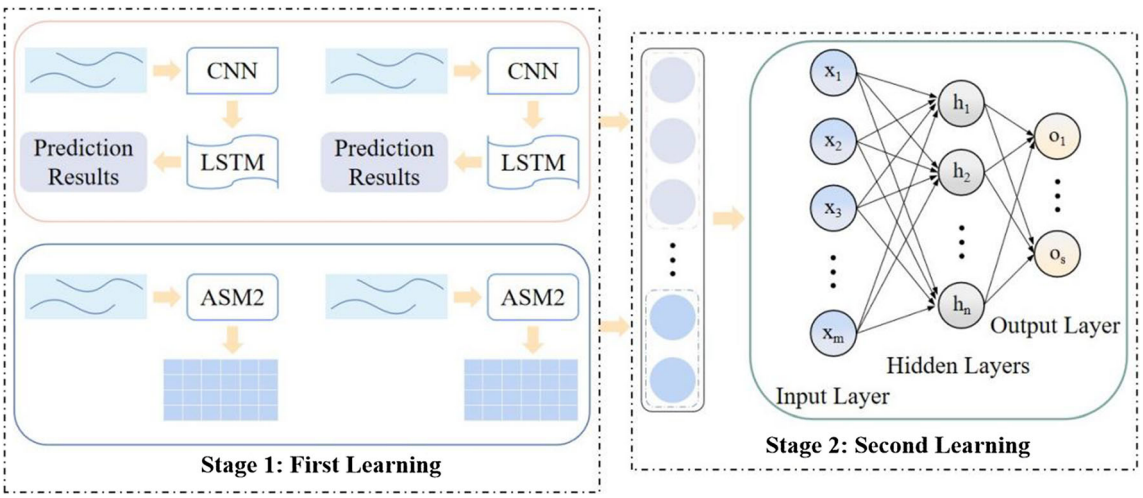


Fig. 2 Framework of the AS-CL mechanism

Table 1 The nomenclature section with acronyms

Acronyms	Definition
WTP	Wastewater Treatment Processes
CNN	Convolutional Neural Network model
LSTM	Long Short-Term Memory model
ASM	Activated sludge models
AS-CL	fuse of Activated Sludge model, Convolutional neural network and Long short-term memory neural networks
DO	Dissolved Oxygen
COD	Chemical Oxygen Demand
NH_4-N	Ammonia nitrogen
A-A-O	Anaerobic-Anoxic–Oxic biological treatment processes
A-A-A-O	Anoxic-Anaerobic-Anoxic–Oxic biological treatment processes

attribute which is the correlations between current water quality and past inputs. To meet the spatial and time attribute of the water quality index, the hybrid CNN-LSTM model is adopted for the effluent index forecast. Concretely, the CNN was employed to extract the data spatial feature, while the LSTM was applied to extract the temporal features, because current pollutant concentrations are suffered by the historical pollutant concentration. Figure 4 presents the architecture of the involved CNN-LSTM model in detail.

1D-CNN is widely used in sequence data processing because of the nature, which includes an input layer, several 1D convolution layers, a pooling layer, and a fully connected layer. LSTM mainly consists of cell state and gating unit which includes forget gate, input gate and output gate. Meanwhile, tanh function and sigmoid function are introduced to overcome gradient disappearance [44] and control the degree of forgetting or remembering information, respectively. The architecture exploits the CNN as the base structure. Firstly, compressing redundant data aims to utilize its convolutional layer, and then extracting features aims to utilize pooling layer. Subsequently, this paper uses a flat layer to process the data format into the one that conforms to the LSTM requirements. The output of the flat layer is used as input to the LSTM. The data is capable of fully filtered, calculated, and output in LSTM. Finally, the fully connected layer is applied to decode the LSTM output and obtain effluent index concentration value in the next period.

The convolution layer is the core part of the convolution neural network. With the traversal of the convolution kernel in each matrix local position sliding from left to right "Z" glyph, the convolution operation is repeated to obtain a complete feature map. The essence of the convolution operation is to extract data features, which refer to the inner product operation between the convolution kernel and local feature matrix, and can be calculated by formula (1).

$$y_{\mathrm{i}} = P\left(y_{i-1} \odot z_{\mathrm{ij}} + b_{\mathrm{i}}\right) \tag{1}$$

where y_{i} is the local feature matrix of the i–th convolution layer, $P(\cdot)$ is the activation function, and $\odot$ is the inner product operation, z_{ij} represents convolution kernel matrix, and b_{i} is the bias vector of the i–th layer.

Pooling refers to downward sampling on the width and height dimensions without changing the depth dimension. Pooling operation is capable of reducing the computation and extracting salient features through downward sampling. The max-pooling method is exploited in this paper, which selects the feature points maximum in the neighborhood as the final eigenvalue, and it is expressed as follows:

$$h_{\mathrm{i}} = \max\_\mathrm{pooling}\left(x_{\mathrm{ij}}\right) \tag{2}$$

where h_{i} is the output of pooling operation, max_pooling$(\cdot)$ is the pooling method, and x_{ij} is the pooling filter matrix. LSTM is a special RNN in which three additional states are introduced to a single cycle structure compared with the RNN. The detailed process of data processing via LSTM is described as follows:

$$f_{\mathrm{n}} = \gamma(\omega_{\mathrm{f}} \cdot [g_{n-1}, s_{\mathrm{n}}] + b_{\mathrm{f}}) \tag{3}$$

$$i_{\mathrm{n}} = \gamma(\omega_{\mathrm{i}} \cdot [g_{n-1}, s_{\mathrm{n}}] + b_{\mathrm{i}}) \tag{4}$$

$$\tilde{C}_{\mathrm{n}} = \tanh(\omega_{\mathrm{c}} \cdot [g_{n-1}, s_{\mathrm{n}}] + b_{\mathrm{c}}) \tag{5}$$

$$C_{\mathrm{n}} = f_{\mathrm{n}} \cdot C_{n-1} + i_{\mathrm{n}} \cdot \tilde{C}_{\mathrm{n}} \tag{6}$$

$$o_{\mathrm{n}} = \gamma(\omega_{\mathrm{o}} \cdot [g_{n-1}, s_{\mathrm{n}}] + b_{\mathrm{o}}) \tag{7}$$

$$h_{\mathrm{n}} = o_{\mathrm{n}} \cdot \tanh(C_{\mathrm{n}}) \tag{8}$$

where f_{n}, i_{n}, and o_{n} represent forget gate, input gate, and output gate expression at time n, respectively; the forget gate f, which characterizes the forget rate of the cell memory given data input; the input gate i, which determines how much proportion of information input shall merge into the cell memory; and the output gate o, which controls how the cell memory shall influence the node output. γ is the sigmoid function, $\gamma = \frac{1}{1+\exp(-x)}$; ω indicates the weight matrix; g_{n-1} indicates the cell state of the previous LSTM; s_{n} means the input information of the current time, and b means the bias parameter; $\tilde{C}_{\mathrm{n}}$ denotes the candidate value vector generated by tanh layer, whose function is expressed as $\tanh = \frac{\exp(x)-\exp(-x)}{\exp(x)+\exp(-x)}$; C_{n} is updated cell state based on the output of the forget gate, the input gate, and the unit state; and h_{n} is the final output of LSTM.

Each node in the fully connected layer is connected to all nodes in the previous layer to integrate the features

Table 2 The symbols used in the methodology

Symbols	Definition
y_{i}	The local feature matrix of the *i*–th convolution layer
z_{ij}	Convolution kernel matrix
b_{i}	The bias vector of the *i*–th convolution layer
h_{i}	The output of pooling operation
x_{ij}	The pooling filter matrix
f_{n}, i_{n}, o_{n}	Forget gate, input gate, and output gate expression at time n, respectively
γ	The sigmoid function
ω	The weight matrix
b	The bias vector
g_{n-1}	The cell state of the previous LSTM
s_{n}	The input information of the current time
$\tilde{C}_{\mathrm{n}}$	The candidate value vector generated by *tanh* layer
C_{n}	Updated cell state based on the output of the forget gate, the input gate, and the unit state
h_{n}	The final output of LSTM
F	The output matrix of fully connected layer
ω_{F}	The weight matrix of the fully connected layer
θ'	Iterated function
$\mathrm{J}(\theta)$	The loss function
x	Eigenvalue
y	Target value
$h_\theta(\cdot)$	Hypothesis function
j	The parameter exponent
w_{i}, w_{e}	The total amount of material in influent and effluent
w_{r}	The number of substances involved in the reaction process
V	The reactor volume
$\left(\frac{dc_u}{dt}\right)$	Tate equation of u
u	Different material components
Q_i, Q_e	The influent and effluent flow rate
$c_{\mathrm{i,u}}$, $c_{e,u}$	The initial concentration and final concentration of u in the reactor
V_{one}, V_{two}, V_{three},V_{four}	The anaerobic, anoxic, oxic tank, and secondary settler volume, respectively
Q_{s}	Returned activated sludge volume
$c_{\mathrm{s,u}}$	Concentration of u in the returned sludge
Q_{one}, Q_{two}, Q_{three},Q_{four}	The effluent flow rate of anaerobic, anoxic, oxic tank, and secondary settler, respectively
$c_{\mathrm{one,u}}$, $c_{\mathrm{two,u}}$, $c_{\mathrm{three,u}}$,$c_{\mathrm{four,u}}$	Effluent concentration of anaerobic, anoxic, oxic tank, and secondary settler, respectively
Q_{o}	Returned fluid volume
$c_{\mathrm{o,u}}$	A concentration of u in the returned fluid
Q_{p}	Sludge discharge volume
v_{t+1}^{i}	The output of the hidden layer
$\beth_{t+1}^{i}$	The output of the multi-layer perceptron network
$\omega_{\beth}$, $b_{\beth}$	Weight and bias of the output layer, respectively
$\mathcal{T}(\varpi,\tilde{\varpi})$	The cross-entropy cost function
$\varpi\left(\varrho^{(\delta)}\right)$	The probability of the true distribution
$\tilde{\varpi}\left(\varrho^{(-\delta)}\right)$	The probability of the model calculated through data
$\overline{m}_{\mathrm{t}}$, $\overline{v}_{\mathrm{t}}$	Correction value of first-order momentum term and second-order momentum term

extracted from the front-end for data prediction. The fully connected layer can be calculated as follows:

$$F = \beta(\omega_{\mathrm{F}} \cdot h_{\mathrm{n}} + b_{\mathrm{F}}) \tag{9}$$

where F is the output matrix, $\beta(\cdot)$ is the activation function; ω_{F} is the weight matrix of the fully connected layer; h_{n} denotes the inputs of fully connected layer; and b_{F} is the bias parameter.

Overfitting is a common phenomenon in the process of deep neural network training because of proportional imbalance between model parameters and training samples. Stochastic gradient descent is one of the simple and effective techniques for addressing this type of phenomenon where a sample is randomly sampled to update parameters. The eigenvalues are iteratively calculated in the form of random sampling to update the parameters which can be expressed as follows:

$$\theta_{\mathrm{i}}^{'} = \theta_{\mathrm{i}} - \alpha \cdot \frac{\partial}{\partial\theta} J(\theta) \tag{10}$$

$$\frac{\partial}{\partial\theta} J(\theta) = \left(y^{\mathrm{j}} - h_\theta\left(x^{\mathrm{j}}\right)\right) x_{\mathrm{i}}^{\mathrm{j}} \tag{11}$$

where θ represents a key parameter in the function, whose initial value can be randomly assigned; θ' stands for iterated function; α is the learning rate, that is, the step length along the gradient negative direction in the iteration process; $J(\theta)$ is the loss function; x and y denote eigenvalue and target value; $h_\theta(\cdot)$ denotes hypothesis function suitable for the algorithm; and j is the parameter exponent, $j \in (1, m)$.

4.2 Knowledge modeling

At present, the knowledge model of wastewater treatment is mainly based on the equation, combined with the reactor and microbial theory, the matrix degradation, and microbial growth process parameters are described to obtain the complete model. However, on the one hand, because these models are complicated, the biochemical reactions are required to be simplified in the use process. On the other hand, some specific parameters of biochemical reactions

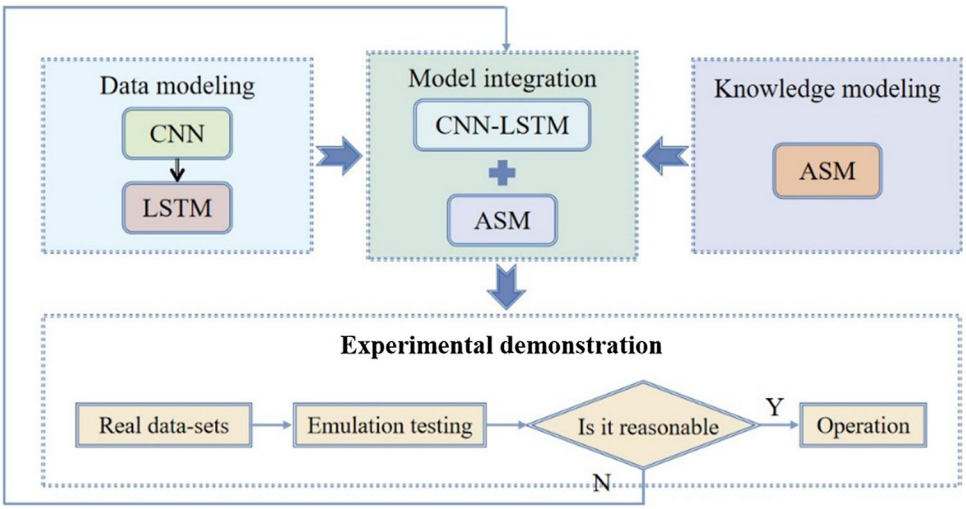


Fig. 3 Technical roadmap of the proposed AS-CL model

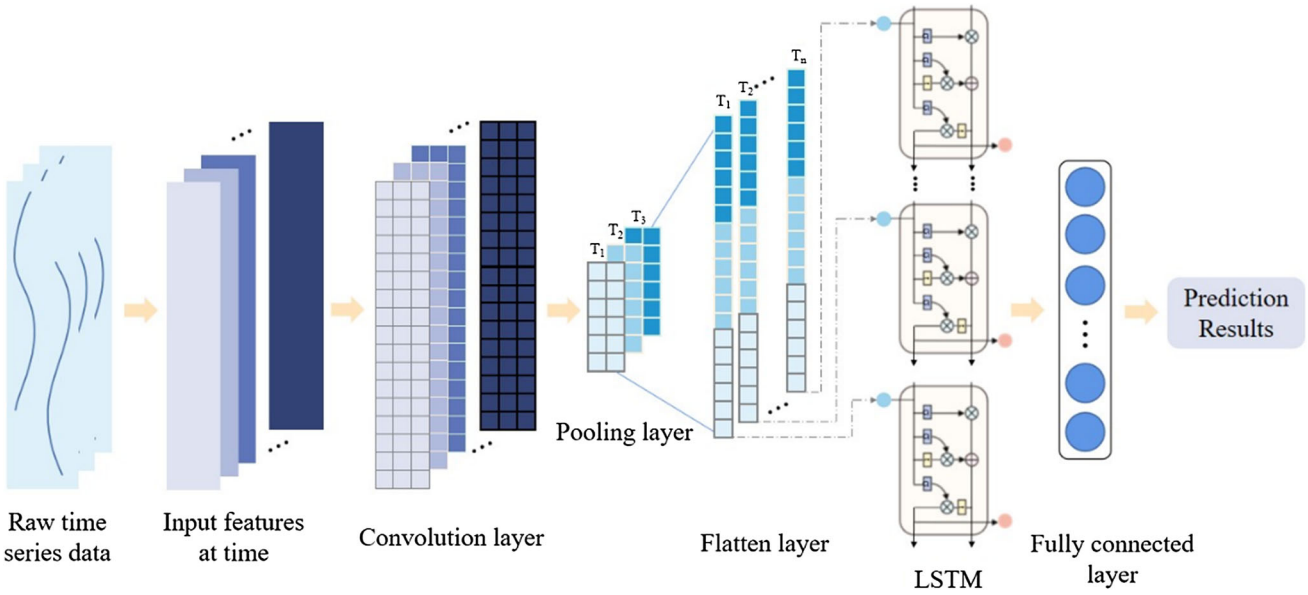


Fig. 4 Flowchart of the CNN-LSTM model

cannot be accurately measured in practice, which hinders the application of the model.

With the research of environmental engineering methods and methods, various methods for the control and treatment of sewage developed with the action of microorganisms as the subject thing have made considerable progress. In particular, activated sludge method, biological membrane method, anaerobic treatment, biological denitrification, and biological phosphorus removal, etc., technologies have been quite mature and increasingly perfect. The activated sludge model (ASM) is a kind of knowledge models which are the most widely used wastewater treatment method at domestic and abroad. The degradation or removal of waste from wastewater is achieved primarily through the breed and maintenance of microbial communities. Taking the Anaerobic-Anoxic-Oxic (A-A-O) of the activated sludge model no.2 (ASM2) model as an example, as shown in Fig. 5, it mainly consists of the aerobic tank, anoxic tank, oxic tank, and secondary clarifiers. In addition, the system also includes an internal circulation system and an external circulation system.

The original ASM2 system involved various parameters which increased the complexity of the model. This study attempts to simplify the model parameters to a certain

extent without affecting the function of the system. According to the material conservation principle, the equilibrium relationship within ASM2 is expressed as follows.

$$w_i + w_r = w_e \tag{12}$$

where w_i and w_e represent the total amount of material in influent and effluent; w_r is the number of substances involved in the reaction process. The specific equation for the change of the concentration of the different substance is explained by following formula:

$$V\left(\frac{dc_u}{dt}\right) = Q_i c_{i,u} - Q_e c_{e,u} + \sum_R V\left(\frac{dc_u}{dt}\right)_R \tag{13}$$

where V is reactor volume; u is different material components; $\left(\frac{dc_u}{dt}\right)$ is rate equation of u; Q_i and Q_e denote the influent and effluent flow rate; $c_{i,u}$ and $c_{e,u}$ express the initial concentration and final concentration of u in the reactor; and $\sum_R \left(\frac{dc_u}{dt}\right)_R$ is the biochemical reaction rate equation of u. To maintain the microbiological communities in the reactor, partial of the phosphorus-containing sludge is re-circulated into the anaerobic tank through external circulation. According to formula (14), the equilibrium equation of this stage is expressed as follows:

$$V_{one}\left(\frac{dc_u}{dt}\right) = Q_i c_{i,u} + Q_s c_{s,u} - Q_{one} c_{one,u} + \sum_R V_{one}\left(\frac{dc_u}{dt}\right)_R \tag{14}$$

where V_{one} represents the anaerobic tank volume; Q_s is returned activated sludge volume; $c_{s,u}$ is concentration of u in the returned sludge; Q_{one} is the effluent flow rate of anaerobic tank; $c_{one,u}$ is effluent concentration of anaerobic tank.

Anoxic tank and oxic tank constitute the internal circulation system of the process, which refers to the circulation process of the mixture from the oxic tank to the anoxic tank. The mixtures provide the microorganisms in the anoxic tank with sufficient carbon source to maintain stable operation, trying to depict the equilibrium relationship between the anoxic tank and oxic tank through the following two formulas:

$$V_{two}\left(\frac{dc_u}{dt}\right) = Q_{one} c_{one,u} + Q_o c_{o,u} - Q_{two} c_{two,u} + \sum_R V_{two}\left(\frac{dc_u}{dt}\right)_R \tag{15}$$

$$V_{three}\left(\frac{dc_u}{dt}\right) = Q_{two} c_{two,u} - Q_o c_{o,u} - Q_{three} c_{three,u} + \sum_R V_{three}\left(\frac{dc_u}{dt}\right)_R \tag{16}$$

where V_{two} and V_{three} denote the anoxic and oxic tank volume; Q_o stands for returned fluid volume; $c_{o,u}$ is a concentration of u in the returned fluid; Q_{two} and Q_{three} represent the effluent flow rate of anoxic tank and oxic tank; and $c_{two,u}$ and $c_{three,u}$ denote the effluent concentration of anoxic tank and oxic tank.

The secondary settler primarily realizes the separation of mud and water via sedimentation. A small number of the sludge return into the anaerobic tank, while the surplus sludge is then recycled or wasted. And the supernatant is discharged as treated water. The process is represented as follows:

$$V_{four}\left(\frac{dc_u}{dt}\right) = Q_{three} c_{three,u} - Q_s c_{s,u} - Q_p c_{p,u} - Q_{four} c_{four,u} + \sum_R V_{four}\left(\frac{dc_u}{dt}\right)_R \tag{17}$$

where V_{four} is the secondary settler volume; Q_p is sludge discharge volume; $c_{p,u}$ equals to $c_{s,u}$; Q_{four} is the effluent flow rate of secondary settler; $c_{four,u}$ is the effluent concentration of secondary settler.

4.3 Model integration

Traditional machine learning is to select the appropriate single model algorithm to obtain the data model with the highest robustness. However, there are some limitations to the improvement of the single model; hence, the method of model fusion in integrated learning emerges at the right moment. The model integration [45, 46] is a method that comprehensively considers different models and integrates their results. The greater different the sub-models are, the better the integration effect is. ASM2d model and CNN-LSTM model improve respective performance from two different aspects of biochemical mechanism characteristics and statistical learning of data characteristics; therefore, the accuracy of prediction results can be improved theoretically by combining the two models.

Figure 2 demonstrates framework of the proposed model integration mechanism. It provides support for wastewater treatment processes management systems by predicting water quality. In this paper, the prediction

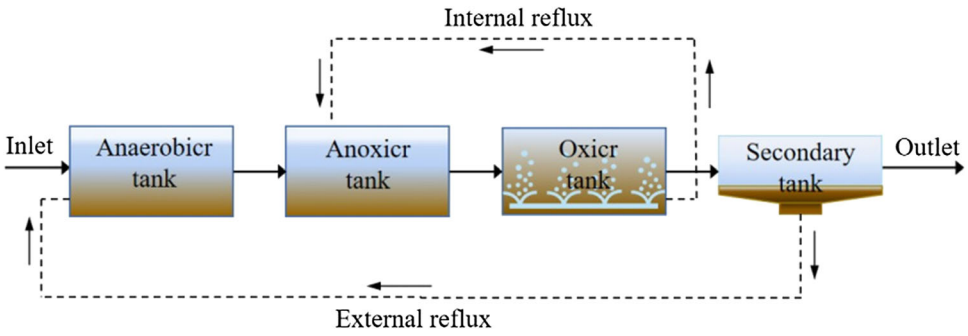


Fig. 5 A-A-O process flowchart

results of the knowledge model and the data model are connected via the full connection layer. The multi-layer perceptron network is utilized in this part, which consists of one input layer, one output layer and two hidden layers. Among them, the first hidden layer has 200 neurons, and the second hidden layer has 160 neurons. The input layer is used to accommodate the integrated data. Moreover the hidden layers and the nonlinear factors are introduced via the activation function ReLU, which makes the neural network more approximate to the real nonlinear relations. Finally, the output layer applies softmax regression to obtain the final prediction results. The prediction results from $h_{\mathrm{t}}^{\mathrm{i}}$ and $w_{\mathrm{e}}^{\mathrm{i}}$ of ASM and CNN-LSTM are integrated, which is represented as follows:

$$\Lambda_{\mathrm{t}}^{\mathrm{i}} = h_{\mathrm{t}}^{\mathrm{i}} \oplus w_{\mathrm{e}}^{\mathrm{i}} \tag{18}$$

The reasonable distribution for weight index can be obtained via the analysis of the influence of different models on the practical management effect, so that to get the predicted value closer to the actual management level. We define the hidden layer function as follows:

$$v_{t+1}^{i} = \varsigma\left(\omega_v \cdot \Lambda_t^i + b_v\right) \tag{19}$$

where v_{t+1}^{i} is the output of the hidden layer; ω_v and b_v denote weight and bias, respectively; $\varsigma(\cdot)$ is the nonlinear activation function, where $\varsigma(\cdot) = \mathrm{ReLU}$. The hidden layer to output layer is similar to a multi-category logistic regression, that is, softmax regression which can be calculated follows:

$$\beth_{t+1}^{i} = \mathrm{softmax}\left(\omega_{\beth} \cdot v_{t+1}^{i} + b_{\beth}\right) \tag{20}$$

where $\beth_{t+1}^{i}$ denotes the output of the multi-layer perceptron network; $\omega_{\beth}$ and $b_{\beth}$ denote weight and bias of the output layer, respectively; and $\mathrm{softmax}(\cdot)$ is the regression function. To measure the suitability of the parameters in the prediction function and improve the accuracy of model parameters, the paper introduced the cross-entropy cost function as optimization objective, which can be represented as follows:

$$\mathcal{T}(\varpi, \tilde{\varpi}) = -\sum_{\delta=1}^{M} \varpi\left(\varrho^{(\delta)}\right) log\tilde{\varpi}\left(\varrho^{(-\delta)}\right) \tag{21}$$

where $\mathcal{T}(\varpi, \tilde{\varpi})$ is used to evaluate the difference between the probability distribution obtained by current training and the real distribution. Reducing cross-entropy loss is to improve the prediction accuracy of the model. $\varpi\left(\varrho^{(\delta)}\right)$ denotes the probability of the true distribution; $\tilde{\varpi}\left(\varrho^{(-\delta)}\right)$ denotes the probability of the model calculated through data; ϱ stands for all parameters in the model integration; and δ denotes a parameter in gradient descent algorithm, which is represented as follows:

$$\delta_{\mathrm{t}} = \delta_{t-1} - \eta \cdot \bar{m}_{\mathrm{t}} / \left(\sqrt{\bar{v}_{\mathrm{t}}} + \epsilon\right) \tag{22}$$

where δ_{t} and δ_{t-1} stand for the parameter of the t_th and $t-1$_th iteration model, respectively; η denotes the learning rate; $\overline{m}_{\mathrm{t}}$ and $\overline{v}_{\mathrm{t}}$ are correction value of first-order momentum term and second-order momentum term, respectively; and ϵ is an arbitrarily small value to avoid the denominator being 0, which is generally $1\mathrm{e}^{-8}$.

The purpose of the optimization is to obtain the parameter set which makes the model more approximate to achieve the optimal value by updating the parameters that affect the model. The optimization algorithm be utilized is the Adam Optimizer. The Adam calculates the adaptive learning rate with different parameters from the estimation for the first and second moments of the gradient to update the network weights. In brief, Adam uses momentum and adaptive learning rate to speed up convergence.

Algorithm: Model Encoding

INPUT: $\Re_{in}, \mathfrak{B}_{in}, \mathfrak{U}_{ij}^{(t)}, y_{(t)}(t = 1,2,\cdots,T)$

OUTPUT: $\Re_{out}, \mathfrak{B}_{out}$

1：While CNN-LSTM do

2： While CNN do

3： the convolution operation y_i as Eq.(1)

4： the max-pooling method h_i as Eq.(2)

5： End while

6： While LSTM do

7： For n=1 to N do

8： forget gate f_n as Eq.(3)

9： Update vector i_n and $\tilde{C}_n$ as Eq.(4) and Eq.(5)

10： input gate C_n as Eq.(6)

11： Update vector o_n as Eq.(7)

12： output gate h_n as Eq.(8)

13： End for

14： End while

15： fully connected layer F as Eq.(9)

16： Update $\theta_i' = \theta_i - \alpha \cdot \frac{\partial}{\partial\theta} J(\theta)$

17：End while

18：While ASM do

19： the equilibrium relationship within ASM2 as Eq.(12)

20： the change the concentration of the different substance as Eq.(13)

21： Define $V_{one}\left(\frac{dc_u}{dt}\right)$、$V_{two}\left(\frac{dc_u}{dt}\right)$、$V_{three}\left(\frac{dc_u}{dt}\right)$ from Eq.(14) to Eq.(16)

22： Define $V_{four}\left(\frac{dc_u}{dt}\right)$ as Eq.(17)

23：End while

24：While AS-CL do

25： For t=1 to T do

26： input layer Λ_t^i as Eq.(18)

27： hidden layer ν_{t+1}^i as Eq.(19)

28： output layer $\beth_{t+1}^i$ as Eq.(20)

29： End for

30： For δ =1 to M do

31： optimization function $\mathcal{T}(\varpi, \widetilde{\varpi})$ as Eq.(21)

32： Update $\delta_t = \delta_{t-1} - \eta \cdot \overline{m}_t / \left(\sqrt{\overline{v}_t} + \epsilon\right)$

33： End for

34：End while

5 Experimental settings

5.1 Data pre-processing

During the data monitoring in the sewage treatment plant, abnormal data phenomenon is frequently caused because of the relevant data records incomplete which refers to monitoring equipment malfunction or failure, and data transmission process error or other factors. In order to obtain a real and effective data-set, it is necessary to pre-process the original monitoring data via a series of scientific means. The specific methods are shown as follows:

- *data consolidation:* consolidating sewage treatment data came from various data sources, such as combining data from monitoring systems and manual records into the same table.
- *data cleaning*: Data loss, data conflict, noise data, and so on are generally found in the data obtained from sewage treatment plants through monitoring systems. Therefore, it is necessary to fill in the missing values, smooth noisy data, identify or remove outliers, as well as to solve inconsistency and others via data cleaning methods.
- *data transformation:* The data is transformed into data form suitable for the modeling process by unit conversion, data normalization, etc. For instance, the random sampling temporal frequency of the monitoring system is inconsistent, while it is necessary to obtain the best data form by averaging the monitoring data in a certain period. Moreover, the data input time unit of the machine model is frequently based on days, however, that of the water quality monitoring frequency is based on the minute, and therefore, the data unit needs to be transformed to make data more consistent with the model.

Figure 6 depicts the parameters and symbols associated with the data-set. DO_A and DO_B stand for DO in different tanks; $\mathfrak{U}$ is the symbolic representation of DO concentration. NH_4-N_i and NH_4-N_o represent influent and effluent NH_4-N, respectively; $\mathfrak{B}$ is the symbolic of NH_4-N concentration. COD_i and COD_o represent influent and effluent COD, respectively; and $\Re$ is the symbol for COD concentration. Statistical features of the parameters are illustrated in Table 3.

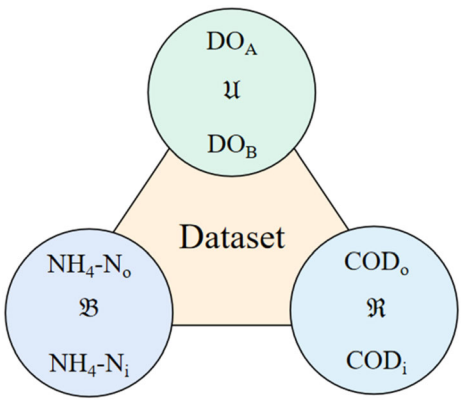


Fig. 6 Symbols of the experimental dataset

Figure 7 contains two sub-figures, which visually reveal the fluctuations of inlet NH_4-N and inlet COD, respectively. Statistical characteristics of DO values in several tanks can be legibly observed from six subfigures of Fig. 8. The DO values in each tank are divided into 5 distribution intervals according to the equal interval. The distribution of dissolved oxygen in each tank can be intuitively distinguished from Fig. 8. Meanwhile, Tank1-A, Tank1-B, Tank2-A, Tank2-B, Tank3-A, Tank3-B separately correspond to variable $\mathfrak{U}_{1,A}$, $\mathfrak{U}_{1,B}$, $\mathfrak{U}_{2,A}$, $\mathfrak{U}_{2,B}$, $\mathfrak{U}_{3,A}$, $\mathfrak{U}_{3,B}$.

5.2 Experimental settings

To quantitatively analyze the performance of the proposed AS-CL, the mean absolute error (MAE) and root-mean-square error (RMSE) are selected to evaluate in this phase, whose formula are represented as follows:

$$\mathrm{MAE} = \frac{1}{\mathcal{J}}\sum_{\tau=1}^{\mathcal{J}}\left|\psi_\tau - \overrightarrow{\psi}_\tau\right| \tag{23}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathcal{J}}\sum_{\tau=1}^{\mathcal{J}}\left(\psi_\tau - \overrightarrow{\psi}_\tau\right)^2} \tag{24}$$

where $\mathcal{I}$ means the test set quantity; $\left(\psi_\tau - \overrightarrow{\psi}_\tau\right)$ indicates the prediction error value. For the above two metrics, the result is inversely proportional to the model performance, that is, the larger the result is, the worse the model performs. In order to verify the stability and reliability of the proposed AS-CL, five methods can be exploited as the

Table 3 Statistical features of the experimental dataset

Variable	Min	Max	Mean	S. D
$\mathfrak{U}_{1,A}$	1.002	9.562	2.810	1.519
$\mathfrak{U}_{1,B}$	1.000	9.287	3.293	1.861
$\mathfrak{U}_{2,A}$	1.001	9.413	2.607	1.117
$\mathfrak{U}_{2,B}$	1.000	9.088	2.691	1.109
$\mathfrak{U}_{3,A}$	1.003	9.956	5.571	2.604
$\mathfrak{U}_{3,B}$	1.241	9.973	6.152	3.118
$\mathfrak{B}_{in}$	15.009	49.501	26.938	5.901
$\mathfrak{B}_{out}$	0.081	0.582	0.165	0.109
$\Re_{in}$	102.619	979.864	518.699	199.750
$\Re_{out}$	19.653	212.127	51.237	49.998

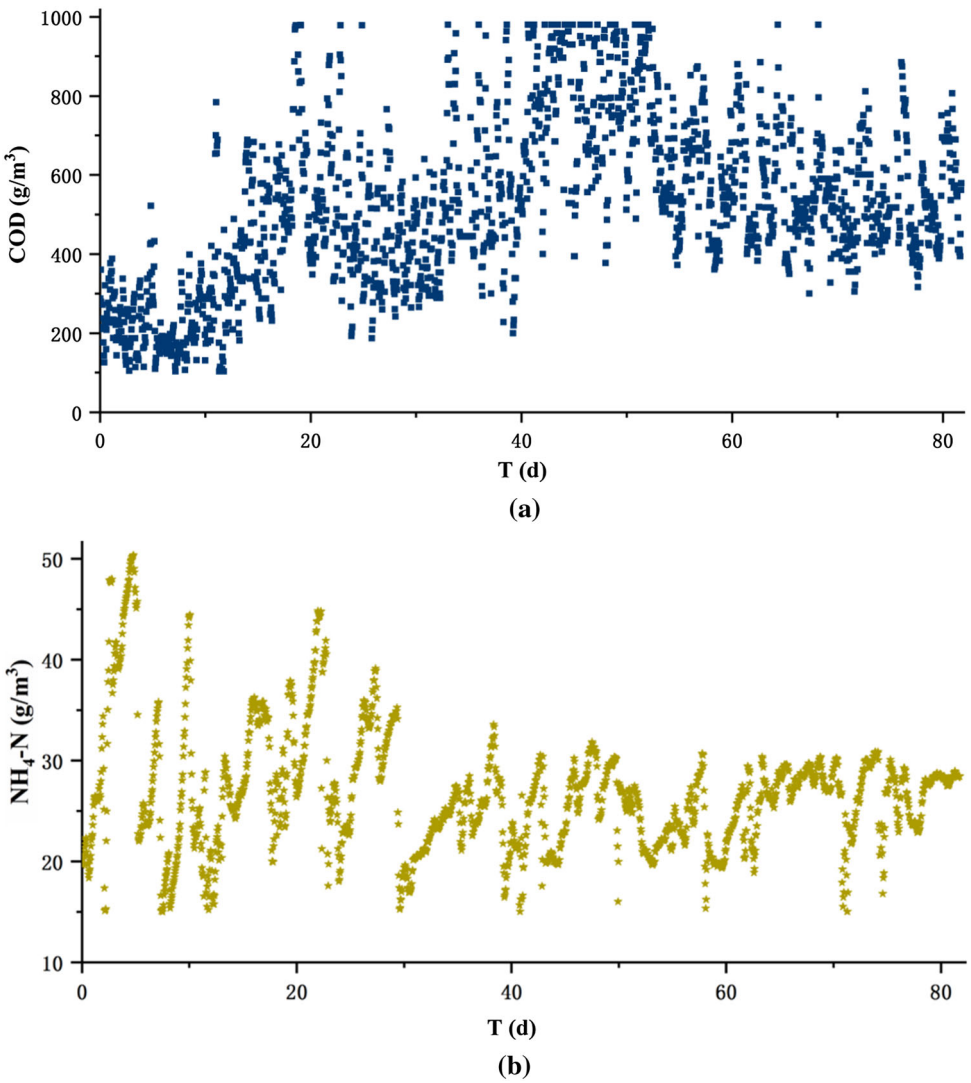


Fig. 7 The fluctuations of inlet COD and inlet NH_4-N

benchmark methods for AS-CL performance evaluation. Comparing the proposed AS-CL with the baseline models during the experiment, the baseline model is introduced as follows and summary is described in Table 4:

- *Multi-layer Preceptor (MLP):* It is a relatively complex artificial neural network (ANN), which has multiple hidden layers in between besides the input and output layer.
- *Convolutional neural network (CNN):* It is a specially deep neural network model, which is mainly reflected in two aspects. For one thing, the connections between its neurons are not fully connected; for another, the weights of connections between certain neurons in the same layer are shared.
- *Long short-term memory (LSTM):* It is an improved recurrent neural network, which superinduces unit states based on the RNN to preserve the long-term state.
- *CNN-LSTM:* It is a hybrid neural network model combining CNN and LSTM, which utilizes CNN to extract spatial features of data and exploits LSTM to extract temporal features.
- *Activated sludge model (ASM):* It realizes the degradation or removal of waste from wastewater through the breed and maintenance of microbial communities.

Adam is used as the model optimizer in this experiment, and the default learning rate (η) is 0.005. During the experiment, the data set was randomly divided into the training set and testing set, among which the default ratio of the training set to the testing set is 7:3.

5.3 Results and analysis

This research conducted a series of experiments to evaluate the performance of the proposed AS-CL from a multi-dimensional perspective. Among, the experimental process

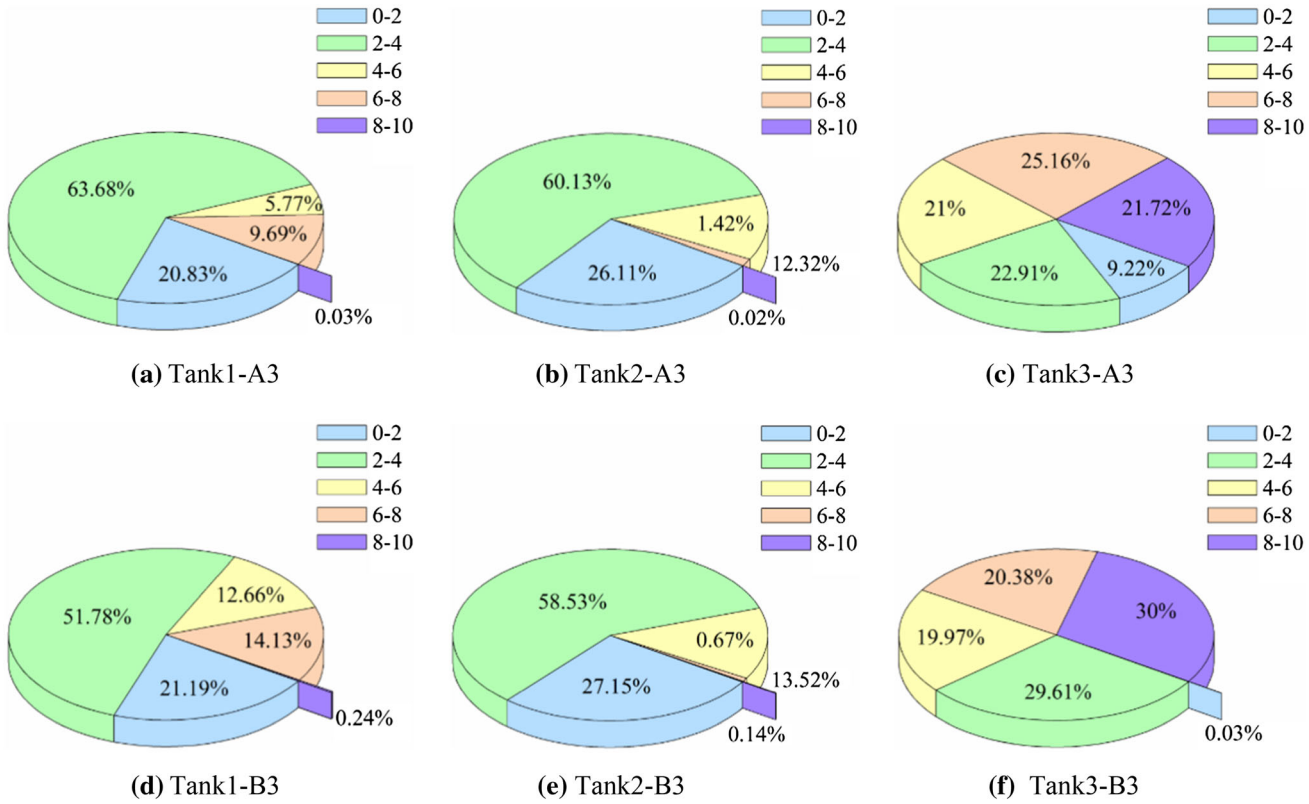


Fig. 8 Statistical characteristics of DO density values distribution in six oxic tanks

Table 4 Baseline model summary

Name	Description	Literature
MLP	A feedforward artificial neural networks in which every neuron is fully connected	[47, 48]
CNN	A kind of feedforward neural network with deep structure that includes convolution calculation	[32, 41]
LSTM	A kind of sequential neural network model specially designed to solve the long-term dependence problem	[49–51]
CNN-LSTM	A hybrid neural network model combining CNN and LSTM in order to capture more complete features	[38, 52, 53]
ASM	A kind of knowledge models which are the most widely used in WTP	[21–23]

of the knowledge model part is achieved by WEST software, and the main setting of the process is mentioned in Sect. 2.1.

With the learning rate and parameter values setting to default values, the training set values are set to 50, 60, 70, and 80%, successively. Tables 5 and 6 exhibit experimental results of AS-CL and baselines under different the training sets, what demands to be clarified is that the bold part in the table represents the minimum average value of each column evaluation index. From the overall trend analysis of experimental data, almost all baseline methods perform better with the increasing proportion of training sets. When the proportion of the training set reaches 70%, the upward trend tends to be saturated. The model performance of two data-sets is intuitively illustrated in Figs. 9 and 10, where X-axis denotes percentage of training set and Y-axis denotes values of evaluation metrics. It can be observed from sub-graphs that the performance of MLP is far behind other methods. It can be also clearly discovered in these sub-graphs that the proposed AS-CL is better than the other five baselines models no matter how the training set value changes. These experimental results can be attributed to three aspects of factors. First of all, ASM model introduces the biochemical reaction mechanism of WTP in this research. Secondly, CNN is adopted to deeply extract the spatial characteristics of WTP monitoring data. Lastly, LSTM is exploited to capture temporal characteristics over a long period.

Subsequently, this paper contrasted and assessed the proposed AS-CL with five baseline methods. In this group of experiments, the training set and parameter values are set to default values, and list the model performance (MAE

Table 5 Results of outlet COD under different training sets

Data-set	Training sets	50%		60%		70%		80%	
	Metrics	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
Daily data-sets	MLP	0.573 ± 0.0049	0.831 ± 0.0052	0.491 ± 0.0053	0.737 ± 0.0053	0.447 ± 0.0051	0.701 ± 0.0050	0.482 ± 0.0053	0.723 ± 0.0052
	CNN	0.323 ± 0.0056	0.412 ± 0.0051	0.277 ± 0.0056	0.383 ± 0.0050	0.256 ± 0.0055	0.362 ± 0.0054	0.268 ± 0.0053	0.374 ± 0.0052
	LSTM	0.242 ± 0.0053	0.337 ± 0.0053	0.187 ± 0.0052	0.297 ± 0.0053	0.151 ± 0.0052	0.239 ± 0.0051	0.174 ± 0.0050	0.285 ± 0.0055
	CNN-LSTM	0.181 ± 0.0056	0.279 ± 0.0053	0.153 ± 0.0049	0.236 ± 0.0051	0.123 ± 0.0050	0.185 ± 0.0051	0.148 ± 0.0050	0.217 ± 0.0051
	ASM	0.163 ± 0.0055	0.234 ± 0.0056	0.125 ± 0.0051	0.193 ± 0.0053	0.109 ± 0.0056	0.166 ± 0.0053	0.114 ± 0.0054	0.187 ± 0.0050
	AS-CL	**0.094** ± 0.0054	**0.173** ± 0.0052	**0.087** ± 0.0052	**0.146** ± 0.0054	**0.081** ± 0.0051	**0.129** ± 0.0052	**0.085** ± 0.0054	**0.138** ± 0.0050

Table 6 Results of outlet NH_4-N under different training sets

Data-set	Training sets	50%		60%		70%		80%	
	Metrics	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
Daily data-sets	MLP	0.534 ± 0.0050	0.801 ± 0.0053	0.478 ± 0.0049	0.616 ± 0.0053	0.425 ± 0.0054	0.592 ± 0.0055	0.437 ± 0.0053	0.639 ± 0.0050
	CNN	0.297 ± 0.0050	0.392 ± 0.0052	0.261 ± 0.0054	0.359 ± 0.0054	0.242 ± 0.0053	0.334 ± 0.0053	0.254 ± 0.0050	0.348 ± 0.0052
	LSTM	0.199 ± 0.0050	0.299 ± 0.0051	0.132 ± 0.0052	0.247 ± 0.0051	0.122 ± 0.0050	0.223 ± 0.0049	0.128 ± 0.0051	0.232 ± 0.0055
	CNN-LSTM	0.136 ± 0.0053	0.231 ± 0.0054	0.122 ± 0.0052	0.204 ± 0.0049	0.109 ± 0.0049	0.172 ± 0.0050	0.117 ± 0.0051	0.199 ± 0.0054
	ASM	0.121 ± 0.0052	0.198 ± 0.0051	0.106 ± 0.0054	0.183 ± 0.0054	0.089 ± 0.0051	0.149 ± 0.0055	0.101 ± 0.0051	0.171 ± 0.0050
	AS-CL	**0.091** ± 0.0054	**0.166** ± 0.0056	**0.086** ± 0.0050	**0.141** ± 0.0050	**0.078** ± 0.0054	**0.121** ± 0.0049	**0.083** ± 0.0053	**0.131** ± 0.0054

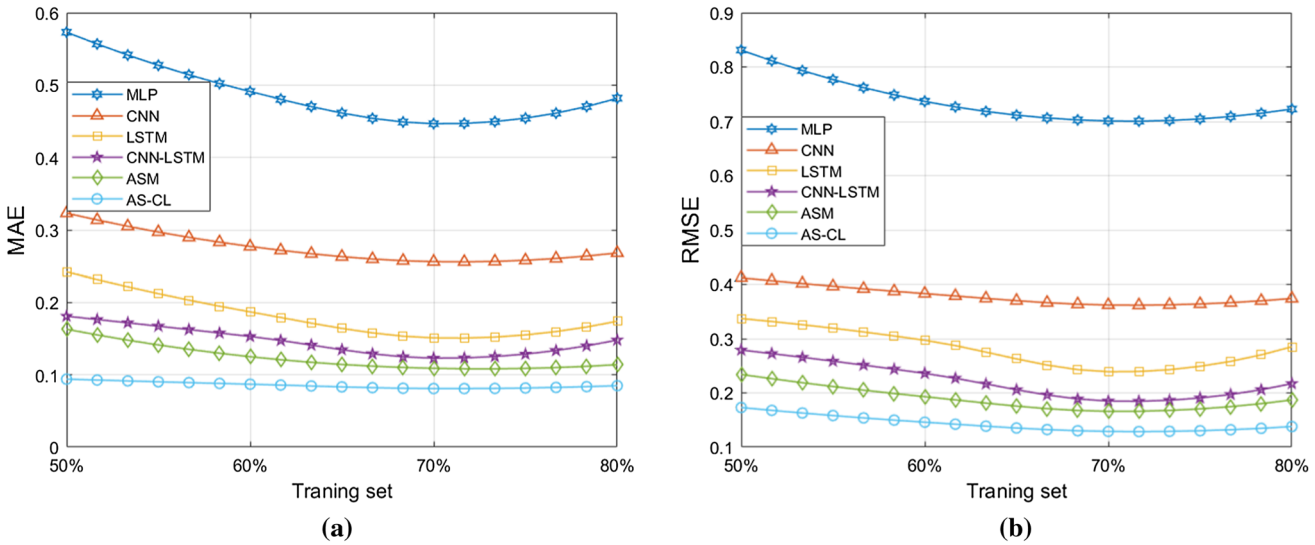


Fig. 9 Results of outlet COD under different values of training set: **a** MAE, **b** RMSE

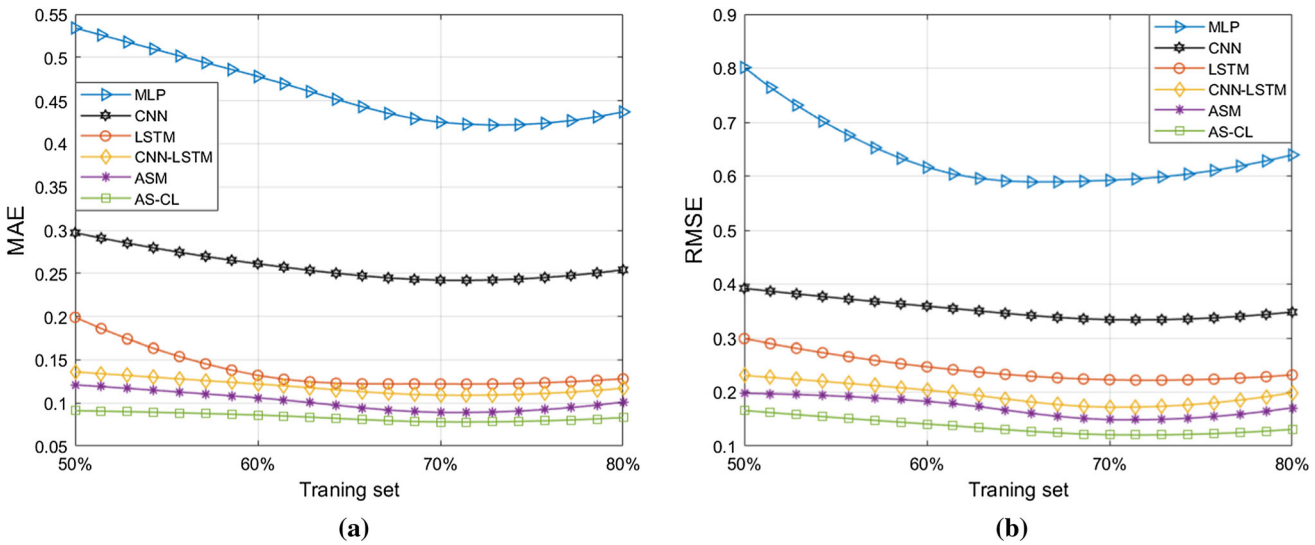


Fig. 10 Results of outlet NH_4-N under different values of training set: **a** MAE, **b** RMSE

and RMSE) of AS-CL and the baseline under the learning rate values set to 0.01, 0.008, 0.005, and 0.003. Tables 7 and 8 clearly display the experimental evaluation results for outlet COD and outlet NH_4-N with different learning rate values, respectively. It can be observed from the two tables that the proposed AS-CL throughout performs better than those of baselines under various learning rates. Figures 11 and 12 more intuitively illustrate the model performance of the two data-sets. Each of them has two subfigures, where X-axis denotes learning rates and Y-axis denotes values of evaluation metrics. It is worth mentioning that the curve is closer to the bottom, which means the smaller the MAE and RMSE value is, the better the model performs. The prediction results of the outlet COD and outlet NH_4-N in the two figures demonstrate that the values of the evaluation indicators MAE and RSME are the minimum while the learning rate is 0.005.

The thirdly set of experiments is implemented to analyze the parameter sensitivity of the proposed AS-CL. In this set of experiments, AS-CL is not compared with any baseline model, nevertheless, prediction results of AS-CL are compared with those of itself after changing the external parameters. Figures 13, 14 and 15, respectively, present the MAE results and RMSE results of AS-CL for predicting outlet COD under different parameter situations. The MAE and RMSE results of AS-CL for outlet NH_4-N under different parameters are, respectively, demonstrated in Figs. 16, 17 and 18. Each data-set contains three figures, corresponding to three different types of parameter combinations: change of the learning rate and training set

Table 7 Results of outlet COD under different learning rates

Data-set	Learning rate	0.003		0.005		0.008		0.01	
	Metrics	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
Daily data-sets	MLP	0.529 ± 0.0054	0.712 ± 0.0053	0.447 ± 0.0051	0.701 ± 0.0055	0.534 ± 0.0051	0.721 ± 0.0055	0.555 ± 0.0053	0.744 ± 0.0051
	CNN	0.265 ± 0.0051	0.377 ± 0.0051	0.256 ± 0.0052	0.362 ± 0.0056	0.273 ± 0.0054	0.380 ± 0.0050	0.284 ± 0.0051	0.386 ± 0.0051
	LSTM	0.164 ± 0.0050	0.285 ± 0.0053	0.151 ± 0.0055	0.239 ± 0.0052	0.170 ± 0.0056	0.294 ± 0.0055	0.182 ± 0.0055	0.324 ± 0.0054
	CNN-LSTM	0.138 ± 0.0053	0.203 ± 0.0051	0.123 ± 0.0052	0.185 ± 0.0055	0.141 ± 0.0053	0.213 ± 0.0054	0.148 ± 0.0052	0.237 ± 0.0052
	ASM	0.112 ± 0.0053	0.183 ± 0.0050	0.109 ± 0.0056	0.166 ± 0.0050	0.117 ± 0.0057	0.189 ± 0.0053	0.129 ± 0.0051	0.198 ± 0.0054
	AS-CL	**0.088** ± 0.0056	**0.143** ± 0.0054	**0.081** ± 0.0054	**0.129** ± 0.0053	**0.089** ± 0.0050	**0.147** ± 0.0051	**0.091** ± 0.0052	**0.154** ± 0.0055

Table 8 Results of outlet NH_4-N under different learning rates

Data-set	Learning rate	0.003		0.005		0.008		0.01	
	Metrics	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
Daily data-sets	MLP	0.523 ± 0.0053	0.607 ± 0.0052	0.425 ± 0.0050	0.592 ± 0.0053	0.529 ± 0.0055	0.609 ± 0.0053	0.539 ± 0.0054	0.613 ± 0.0054
	CNN	0.248 ± 0.0051	0.351 ± 0.0051	0.242 ± 0.0050	0.334 ± 0.0053	0.256 ± 0.0053	0.358 ± 0.0052	0.276 ± 0.0052	0.363 ± 0.0051
	LSTM	0.129 ± 0.0053	0.229 ± 0.0051	0.122 ± 0.0050	0.223 ± 0.0051	0.132 ± 0.0052	0.232 ± 0.0052	0.137 ± 0.0052	0.236 ± 0.0054
	CNN-LSTM	0.113 ± 0.0052	0.197 ± 0.0054	0.109 ± 0.0054	0.172 ± 0.0050	0.115 ± 0.0053	0.200 ± 0.0050	0.118 ± 0.0054	0.204 ± 0.0053
	ASM	0.101 ± 0.0051	0.164 ± 0.0050	0.089 ± 0.0052	0.149 ± 0.0055	0.102 ± 0.0054	0.169 ± 0.0055	0.103 ± 0.0055	0.178 ± 0.0053
	AS-CL	**0.083** ± 0.0055	**0.126** ± 0.0054	**0.078** ± 0.0054	**0.121** ± 0.0050	**0.084** ± 0.0054	**0.131** ± 0.0051	**0.086** ± 0.0052	**0.135** ± 0.0052

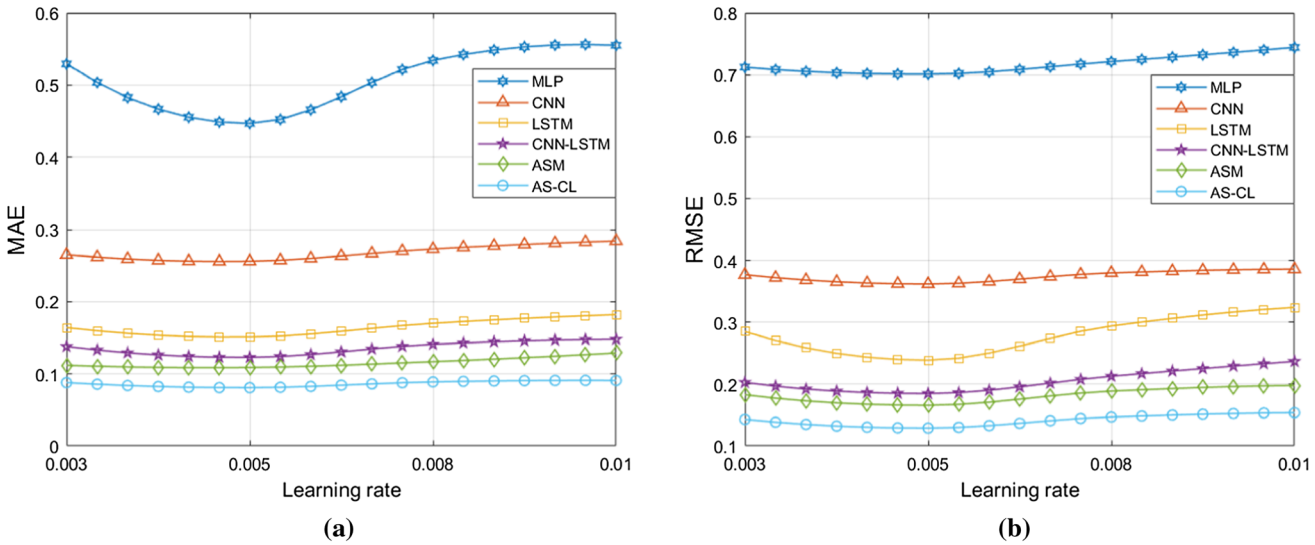


Fig. 11 Results of outlet COD under different learning rates: **a** MAE, **b** RMSE

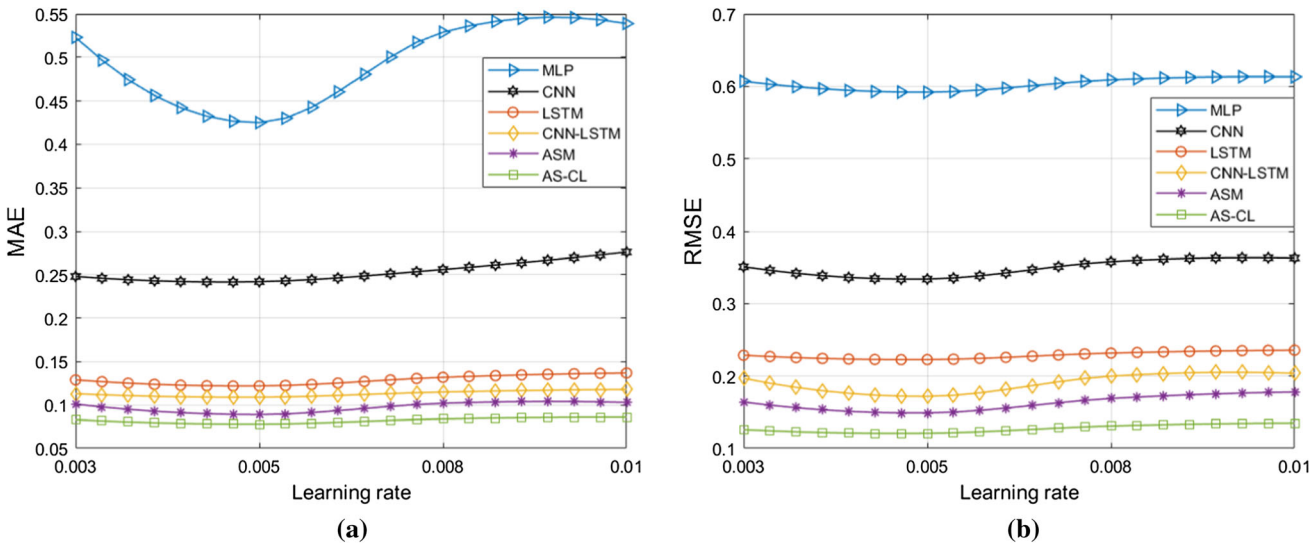


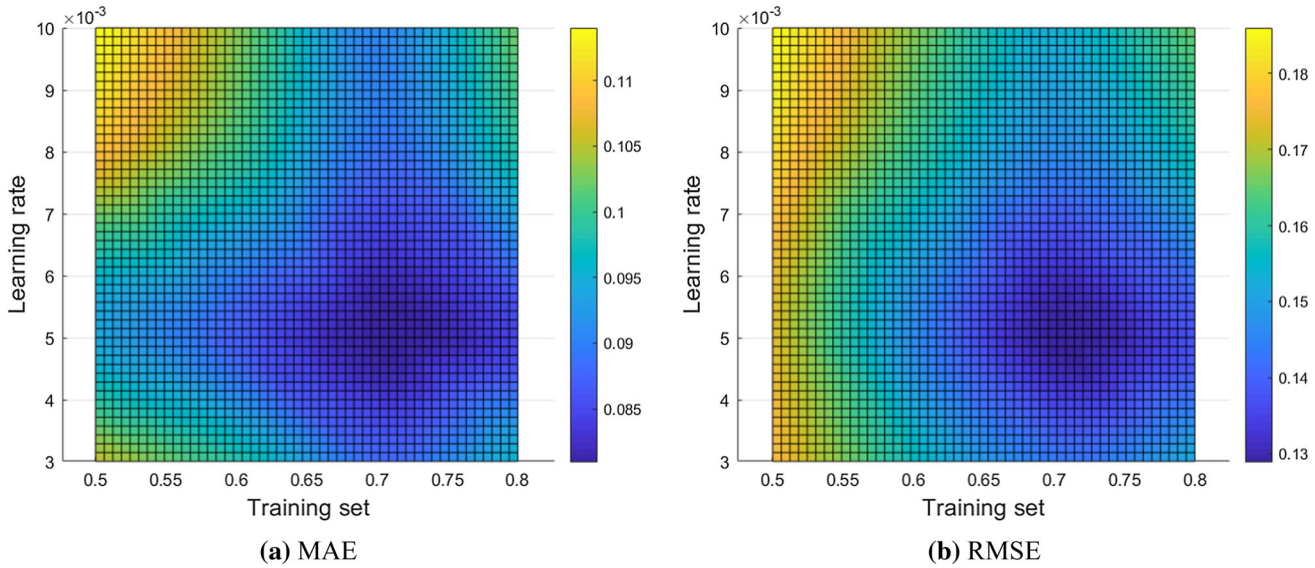
Fig. 12 Results of outlet NH_4-N under different learning rates: **a** MAE, **b** RMSE

proportion, change of the learning rate proportion and optimize category, and change of the training set proportion and optimizer category. It can be intuitively found from the twelve sub-figures that the experimental results of each set fluctuate slightly and remain relatively stable under the condition of different parameter combinations. From these experimental results, it can be concluded that AS-CL captures the complete characteristics of WTP from different perspectives, which makes itself less affected by parameters change. Therefore, it proves that the proposed AS-CL has reasonable stability.

To sum up, the above multi-sets of experiments verified that the proposed AS-CL has superior reliability and stability.

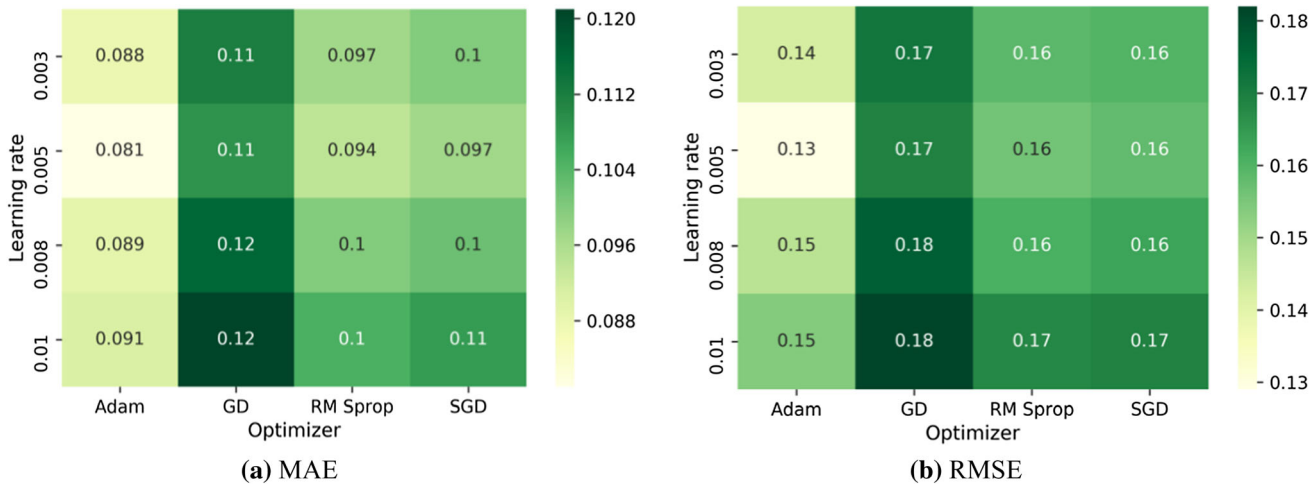
6 Conclusions

WTP is an indispensable part of the urban water circulation system that treats pollutants in wastewater to meet the requirements of different usage. However, traditional sewage treatment technology has become the major hidden trouble that restricts the improvement of sewage treatment efficiency. Currently, WTP modeling is a hot topic



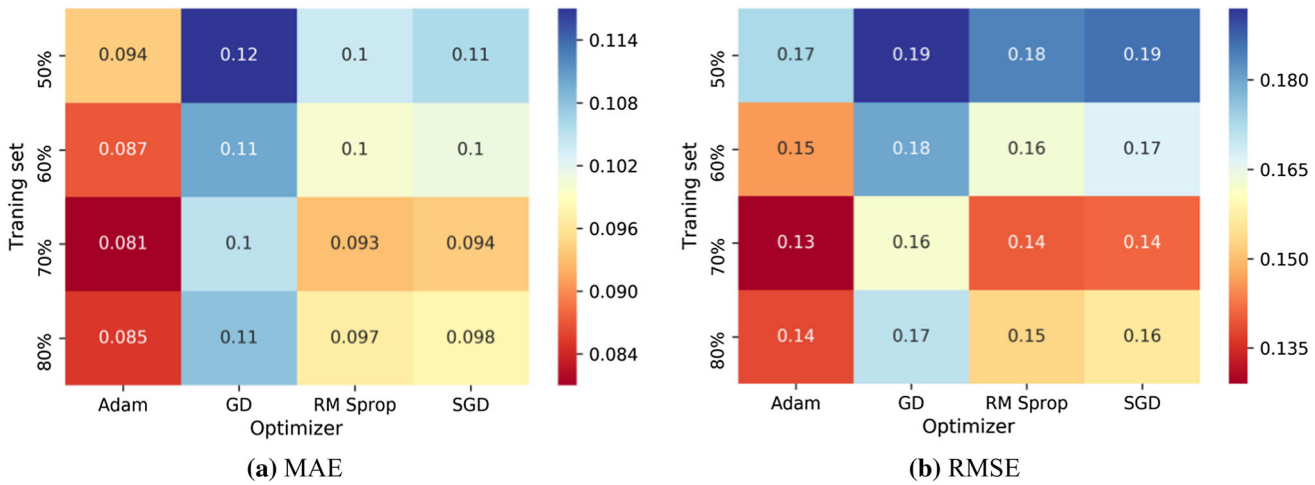
(a) MAE **(b)** RMSE

Fig. 13 The learning rate and training set proportion results of AS-CL with respect to outlet COD under different evaluation metrics



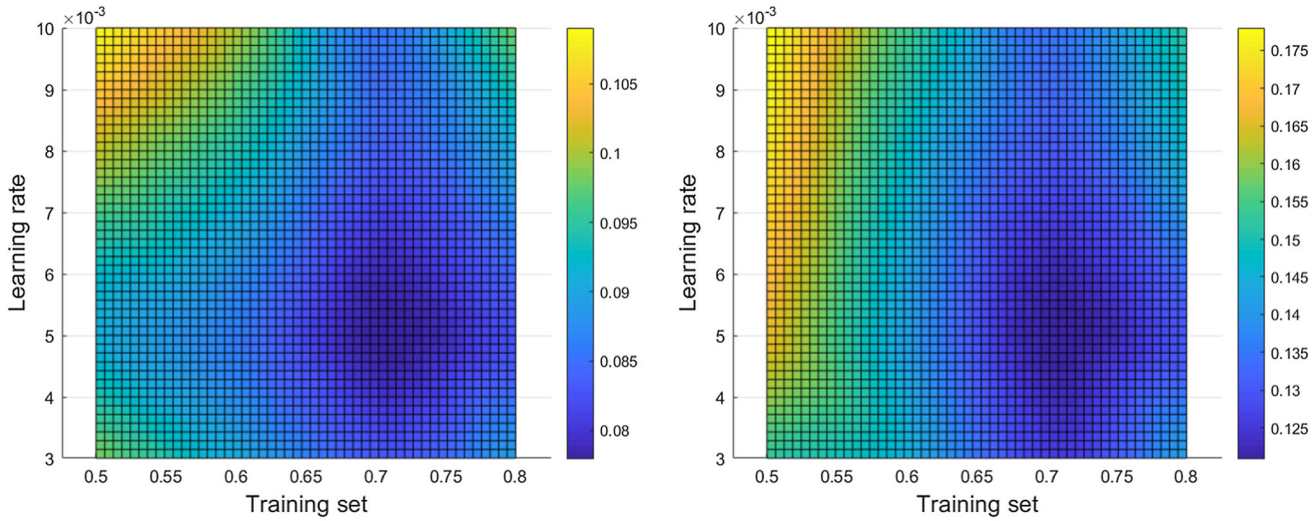
(a) MAE **(b)** RMSE

Fig. 14 The learning rate proportion and optimize category results of AS-CL with respect to outlet COD under different evaluation metrics



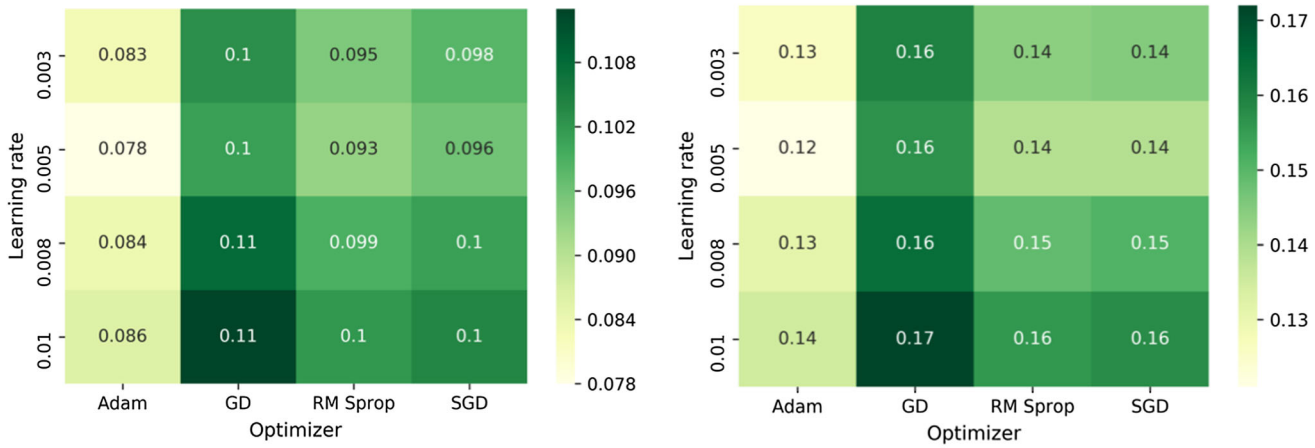
(a) MAE **(b)** RMSE

Fig. 15 The training set proportion and optimizer category results of AS-CL with respect to outlet COD under different evaluation metrics



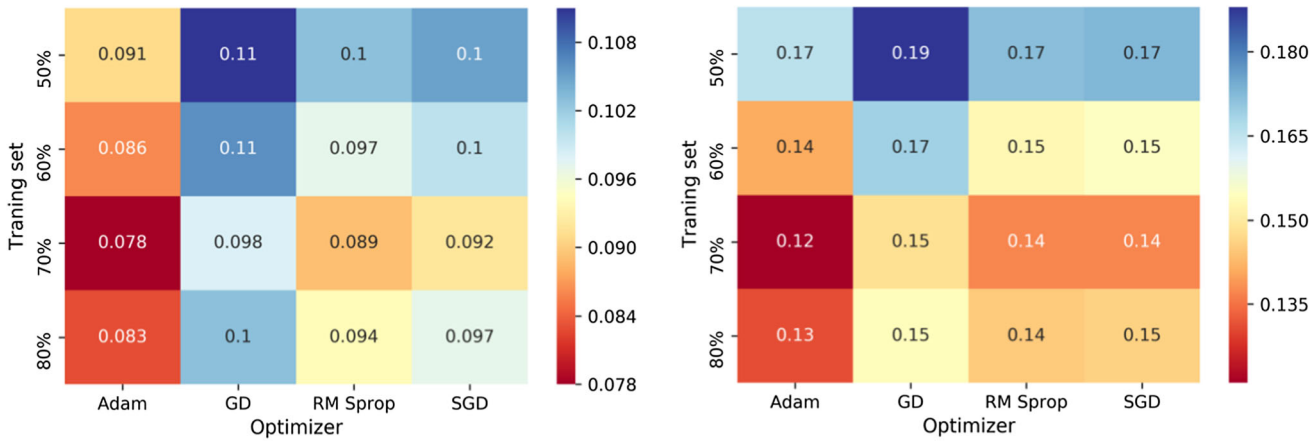
(a) MAE **(b)** RMSE

Fig. 16 The learning rate and training set proportion results of AS-CL with respect to outlet NH_4-N under different evaluation metrics



(a) MAE **(b)** RMSE

Fig. 17 The learning rate proportion and optimize category results of AS-CL with respect to outlet NH_4-N under different evaluation metrics



(a) MAE **(b)** RMSE

Fig. 18 The training set proportion and optimizer category results of AS-CL with respect to outlet NH_4-N under different evaluation metrics

discussed by many scholars. Although significant progress has been obtained, almost all the existing researches are based on one knowledge model or one data model. The former is eminently dependent on the biochemical mechanism, ignoring the hidden features of numerous monitoring data in the WTP, while the latter merely focuses on mining data features, ignoring the impact of biochemical reaction mechanism on WTP. To handle this challenge, this research introduces an AS-CL hybrid system for WTP.

The research carries out three sets of experiments to demonstrate the performance of the proposed AS-CL. To begin with, this paper adjusted the values of training set under the condition of giving the default learning rate and optimizer. Secondly, the learning rate under the condition of giving the default values of training set and optimizer were also adjusted. It should be explained that the first two sets of experiments require to compare AS-CL with the baseline model. The experimental results indicate that the proposed AS-CL performs better than baselines under different parameter settings. Finally, the parameter sensitivity analysis is implemented independently under the combination of different parameters. The experimental results express that the results of each set hardly change under different parameter combinations, proving the stability of the proposed AS-CL.

Acknowledgements This research was supported by National Key Research and Development Program of China (2016YFE0205600), Science and Technology Research Project of Chongqing Municipal Education Commission (KJZD-M202000801), Natural Science Foundation of Chongqing Science & Technology Commission (cstc2020jcyj-msxmX0721), and Science and Technology Research Program of Chongqing Municipal Education Commission (KJQN202000810), Project of Chongqing Technology and Business University (ZDPTTD201917, KFJJ2019053), and Japan Society for the Promotion of Science (JSPS) Grants-in-Aid for Scientific Research (KAKENHI) under Grant JP18K18044.

In addition, we would like to thank the lecturer Zhiwei Guo from Chongqing Technology and Business University, as he had given a number of professional comments during the process of writing and revision. And we also would like to thank the Engineer Dong Feng from Chongqing Sino French Environmental Excellence Research & Development Center Co. Ltd., as he provided experimental datasets from real-world wastewater treatment plants for evaluation.

Declarations

Conflict of interest There are no conflicts of interest to state.

References

1. Ansari M, Othman F, El-Shafie A (2020) Optimized fuzzy inference system to enhance prediction accuracy for influent characteristics of a sewage treatment plant. Sci Total Environ 722:137878. https://doi.org/10.1016/j.scitotenv.2020.137878
2. Chiu SLH, Lo IMC (2018) Identifying key process parameters for uncertainty propagation in environmental life cycle assessment for sewage sludge and food waste treatment. J Clean Prod 174:966–976. https://doi.org/10.1016/j.jclepro.2017.10.164
3. Liu J, Huang L, Buyukada M, Evrendilek F (2017) Response surface optimization, modeling and uncertainty analysis of mass loss response of co-combustion of sewage sludge and water hyacinth. Appl Therm Eng 125:328–335. https://doi.org/10.1016/j.applthermaleng.2017.07.008
4. Zhen L, Zhang Y et al (2021) Early collision detection for massive random access in satellite-based Internet of Things. IEEE Trans Veh Technol 70(5):5184–5189. https://doi.org/10.1109/TVT.2021.3076015
5. Guo Z, Yu K et al (2021) Deep learning-embedded social internet of things for ambiguity-aware social recommendations. IEEE Trans Netw Sci Eng. https://doi.org/10.1109/TNSE.2021.3049262
6. Khatri N, Khatri KK, Sharma A (2020) Enhanced energy saving in wastewater treatment plant using dissolved oxygen control and hydrocyclone. Environ Technol Innov 18:100678. https://doi.org/10.1016/j.eti.2020.100678
7. Huang F, Shen W, Zhang X (2020) Panagiotis Seferlis, Impacts of dissolved oxygen control on different greenhouse gas emission sources in wastewater treatment process. J Clean Prod 274:123233. https://doi.org/10.1016/j.jclepro.2020.123233
8. Skouteris G, Rodriguez-Garcia G, Reinecke SF, Hampel U (2020) The use of pure oxygen for aeration in aerobic wastewater treatment: a review of its potential and limitations. Bioresour Technol 312:123595. https://doi.org/10.1016/j.biortech.2020.123595
9. Xuan Do T, Prajitno H, Il LY, Kim J (2019) Process modeling and economic analysis for bio-heavy-oil production from sewage sludge using supercritical ethanol and methanol. J Supercrit Fluids 150:137–146. https://doi.org/10.1016/j.supflu.2019.05.001
10. Meng D, Xiao Y et al (2021) A data-driven intelligent planning model for UAVs routing networks in mobile Internet of Things. Comput Commun 179:231–241. https://doi.org/10.1016/j.comcom.2021.08.014
11. Man Y, Shen W, Chen X et al (2017) Modeling and simulation of the industrial sequencing batch reactor wastewater treatment process for cleaner production in pulp and paper mills. J Clean Prod 167:643–652. https://doi.org/10.1016/j.jclepro.2017.08.236
12. Zhao L, Dai T, Qiao Z et al (2020) Application of artificial intelligence to wastewater treatment: a bibliometric analysis and systematic review of technology, economy, management, and wastewater reuse. Process Saf Environ Prot 133:169–182. https://doi.org/10.1016/j.psep.2019.11.014
13. Duan H, Gao S, Li X et al (2020) Improving wastewater management using free nitrous acid (FNA). Water Res 171:115382. https://doi.org/10.1016/j.watres.2019.115382
14. Gómez-Llanos E, Durán-Barroso P, Matías-Sánchez A (2018) Management effectiveness assessment in wastewater treatment plants through a new water footprint indicator. J Clean Prod 198:463–471. https://doi.org/10.1016/j.jclepro.2018.07.062
15. Yan W, Xu R, Wang K et al (2020) Soft sensor modeling method based on semisupervised deep learning and its application to wastewater treatment plant. Ind Eng Chem Res 59:4589–4601. https://doi.org/10.1021/acs.iecr.9b05087
16. Vilardi G, Bavasso I, Scarsella M et al (2020) Fenton oxidation of primary municipal wastewater treatment plant sludge: Process modelling and reactor scale-up. Process Saf Environ Prot 140:46–59. https://doi.org/10.1016/j.psep.2020.05.002
17. Sakiewicz P, Piotrowski K, Ober J et al (2020) Optimizing the inclined plate settler for a high-rate microaerobic activated sludge process for domestic wastewater treatment: a theoretical model and experimental validation. Int Biodeterior Biodegrad 154:105060. https://doi.org/10.1016/j.ibiod.2020.105060

18. Wu X, Yang Y, Wu G et al (2016) Simulation and optimization of a coking wastewater biological treatment process by activated sludge models (ASM). J Environ Manag 165:235–242. https://doi.org/10.1016/j.jenvman.2015.09.041
19. Sun Y, Liu J et al (2021) PMRSS: privacy-preserving medical record searching scheme for intelligent diagnosis in IoT healthcare. IEEE Trans Ind Inform. https://doi.org/10.1109/TII.2021.3070544
20. Sakiewicz P, Piotrowski K, Ober J et al (2020) Innovative artificial neural network approach for integrated biogas – wastewater treatment system modelling: effect of plant operating parameters on process intensification. Renew Sustain Energy Rev 124:109784. https://doi.org/10.1016/j.rser.2020.109784
21. Lindow F, Muñoz C, Jaramillo F et al (2020) Active biomass estimation based on ASM1 and on-line OUR measurements for partial nitrification processes in sequencing batch reactors. J Environ Manag 273:111150. https://doi.org/10.1016/j.jenvman.2020.111150
22. Yang SS, Pang JW, Guo WQ et al (2017) Biological phosphorus removal in an extended ASM2 model: Roles of extracellular polymeric substances and kinetic modeling. Biores Technol 232:412–416. https://doi.org/10.1016/j.biortech.2017.01.048
23. Mannina G, Cosenza A, Viviani G, Ekama GA (2018) Sensitivity and uncertainty analysis of an integrated ASM2d MBR model for wastewater treatment. Chem Eng J 351:579–588. https://doi.org/10.1016/j.cej.2018.06.126
24. Chen W, Dai H, Han T et al (2020) Mathematical modeling and modification of a cycle operating activated sludge process via the multi-objective optimization method. J Environ Chem Eng 8:104470. https://doi.org/10.1016/j.jece.2020.104470
25. Karlikanovaite-Balikci A, Yagci N (2019) Determination and evaluation of kinetic parameters of activated sludge biomass from a sludge reduction system treating real sewage by respirometry testing. J Environ Manag 240:303–310. https://doi.org/10.1016/j.jenvman.2019.03.131
26. Harrou F, Dairi A, Sun Y, Senouci M (2018) Wastewater treatment plant monitoring via a deep learning approach. In: 2018 IEEE international conference on industrial technology (ICIT) pp 1544–1548 https://doi.org/10.1109/ICIT.2018.8352410
27. Li H, Chen X et al (2021) Data-driven peer-to-peer blockchain framework for water consumption management. Peer-to-Peer Netw Appl 14:2887–2900. https://doi.org/10.1007/s12083-021-01121-6
28. Zhou P, Zhang R, Xie J et al (2021) Data-driven monitoring and diagnosing of abnormal furnace conditions in blast furnace ironmaking: an integrated PCA-ICA method. IEEE Trans Ind Electron 68(1):622–631. https://doi.org/10.1109/TIE.2020.2967708
29. Krzywanski J, Grabowska K, Sosnowski M et al (2019) An adaptive neuro-fuzzy model of a re-heat two-stage adsorption chiller. Therm Sci 23(Suppl. 4):1053–1063. https://doi.org/10.2298/TSCI19S4053K
30. Clark MC, Hall LO, Goldgof DB et al (2017) Unsupervised brain tumor segmentation using knowledge-based fuzzy techniques. Fuzzy Neuro Fuzzy Syst Med. https://doi.org/10.1201/9780203713419
31. Zuluaga-Bedoya C, Ruiz-Botero M, Ospina-Alarcón M, Garcia-Tirado J (2018) A dynamical model of an aeration plant for wastewater treatment using a phenomenological based semi-physical modeling methodology. Comput Chem Eng 117:420–432. https://doi.org/10.1016/j.compchemeng.2018.07.008
32. Hu P, Tong J, Wang J et al (2019) A hybrid model based on CNN and Bi-LSTM for urban water demand prediction. 2019 IEEE congress on evolutionary computation (CEC) pp 1088–1094 https://doi.org/10.1109/CEC.2019.8790060
33. Yaqub M, Asif H, Kim S et al (2020) Modeling of a full-scale sewage treatment plant to predict the nutrient removal efficiency using a long short-term memory (LSTM) neural network. J Water Process Eng 37:101388. https://doi.org/10.1016/j.jwpe.2020.101388
34. Wang Z, Man Y, Hu Y et al (2019) A deep learning based dynamic COD prediction model for urban sewage. Environ Sci Water Res Technol 5:2210–2218. https://doi.org/10.1039/c9ew00505f
35. Nasser AA, Rashad MZ, Hussein SE (2020) A two-layer water demand prediction system in urban areas based on micro-services and LSTM neural networks. IEEE Access 8:147647–147661. https://doi.org/10.1109/ACCESS.2020.3015655
36. Spérandio M, Espinosa MC (2008) Modelling an aerobic submerged membrane bioreactor with ASM models on a large range of sludge retention time. Desalination 231:82–90. https://doi.org/10.1016/j.desal.2007.11.040
37. Kim H, Lim H, Wie J et al (2014) Optimization of modified ABA2 process using linearized ASM2 for saving aeration energy. Chem Eng J 251:337–342. https://doi.org/10.1016/j.cej.2014.04.076
38. Li T, Hua M, Wu X (2020) A hybrid CNN-LSTM model for forecasting particulate matter (PM2.5). IEEE Access 8:26933–26940. https://doi.org/10.1109/ACCESS.2020.2971348
39. Guo Z, Shen Y, Bashir 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
40. Yu K, Guo Z et al (2021) Secure artificial intelligence of things for implicit group recommendations. IEEE Internet Things J. https://doi.org/10.1109/JIOT.2021.3079574
41. Ding F, Zhu G et al (2020) Deep-learning-empowered digital forensics for edge consumer electronics in 5G hetNets. IEEE Consumer Electron Mag. https://doi.org/10.32604/cmc.2020.014220
42. Su Y (2018) Measuring knowledge diffusion efficiency in R&D networks. Knowl Manag Res Pract 16:208–219. https://doi.org/10.1080/14778238.2018.1435186
43. Skrobek D, Krzywanski J, Sosnowski M et al (2020) Prediction of sorption processes using the deep learning methods (long short-term memory). Energies 13:1–16. https://doi.org/10.3390/en13246601
44. Guo Z, Yu K, Jolfaei A et al (2021) A fuzzy detection system for rumors through explainable adaptive learning. IEEE Trans Fuzzy Syst 6706:1–16. https://doi.org/10.1109/TFUZZ.2021.3052109
45. Yu K, Tan L et al (2021) deep-learning-empowered breast cancer auxiliary diagnosis for 5GB remote e-health. IEEE Wirel Commun 28(3):54–61. https://doi.org/10.1109/MWC.001.2000374
46. Yu K, Tan L et al (2021) Securing critical infrastructures: deep learning-based threat detection in the IIoT. IEEE Commun Mag. https://doi.org/10.1109/MCOM.101.2001126
47. Zhang D, Stewart E, Ye J et al (2020) Roller bearing degradation assessment based on a deep MLP convolution neural network considering outlier regions. IEEE Trans Instrum Meas 69(6):2996–3004. https://doi.org/10.1109/TIM.2019.2929669
48. Alahi A, Goel K, Ramanathan V et al (2016) Social LSTM: human trajectory prediction in crowded spaces. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 961–971 https://doi.org/10.1109/CVPR.2016.110
49. Guo Z, Shen Y et al (2021) Graph embedding-based intelligent industrial decision for complex sewage treatment processes. Int J Intell Syst. https://doi.org/10.1002/int.22540
50. Di Z, Lin J, Peng Q et al (2018) Modeling and simulating of reservoir operation using the artificial neural network, support vector regression, deep learning algorithm. J Hydrol 565:720–736. https://doi.org/10.1016/j.jhydrol.2018.08.050

51. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735
52. Yan R, Liao J, Yang J et al (2021) Multi-hour and multi-site air quality index forecasting in Beijing using CNN, LSTM, CNN-LSTM, and spatiotemporal clustering. Expert Syst Appl 169:114513. https://doi.org/10.1016/j.eswa.2020.114513
53. Feng C, Liu B et al (2021) Blockchain-empowered decentralized cross-domain federated learning for 5G-enabled UAVs. IEEE Trans Ind Inform. https://doi.org/10.1109/TII.2021.3116132