Contents lists available at ScienceDirect

Water Research X

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

Long short-term memory models of water quality in inland water environments

JongCheol Pyo^a, Yakov Pachepsky^b, Soobin Kim^{c,e}, Ather Abbas^d, Minjeong Kim^e, Yong Sung Kwon^f, Mayzonee Ligaray^g, Kyung Hwa Cho^{h,*}

^a Department for Environmental Engineering, Pusan National University, Busan 46241, Republic of Korea

^b Environmental Microbial and Food Safety Laboratory, USDA-ARS, Beltsville, MD, USA

^c School of Civil, Urban, Earth, and Environmental Engineering, Ulsan National Institute of Science and Technology, 50 UNIST-gil, Ulju-gun, Ulsan 44919, Republic of Korea

^d Physical Sciences and Engineering, King Abdullah University of Science and Technology, Thuwal 23955-6900, Kingdom of Saudi Arabia

^e Disposal Safety Evaluation R&D Division, Korea Atomic Energy Research Institute (KAERI), 111, Daedeok-daero 989 beon-gil, Yuseong-gu, Daejeon 34057, Republic of Korea

^f Environmental Impact Assessment Team, Division of Ecological Assessment Research, National Institute of Ecology, Seocheon, Republic of Korea

^g Institute of Environmental Science and Meteorology, College of Science, University of the Philippines Diliman, Quezon City 1101, Philippines

^h School of Civil, Environmental and Architectural Engineering, Korea University, Seoul 02841, Republic of Korea

ARTICLE INFO

Keywords: Long short-term memory Inland water Water quality Ensemble LSTM Deep learning models

ABSTRACT

Water quality is substantially influenced by a multitude of dynamic and interrelated variables, including climate conditions, landuse and seasonal changes. Deep learning models have demonstrated predictive power of water quality due to the superior ability to automatically learn complex patterns and relationships from variables. Long short-term memory (LSTM), one of deep learning models for water quality prediction, is a type of recurrent neural network that can account for longer-term traits of time-dependent data. It is the most widely applied network used to predict the time series of water quality variables. First, we reviewed applications of a standalone LSTM and discussed its calculation time, prediction accuracy, and good robustness with process-driven numerical models and the other machine learning. This review was expanded into the LSTM model with data pre-processing techniques, including the Complete Ensemble Empirical Mode Decomposition with Adaptive Noise method and Synchrosqueezed Wavelet Transform. The review then focused on the coupling of LSTM with a convolutional neural network, attention network, and transfer learning. The coupled networks demonstrated their performance over the standalone LSTM model. We also emphasized the influence of the static variables in the model and used the transformation method on the dataset. Outlook and further challenges were addressed. The outlook for research and application of LSTM in hydrology concludes the review.

1. Introduction

Water monitoring and management of water quality have provided diverse and numerous datasets that have been facilitated by the realization of the impact of global climate change on water resources and by advances in sensor technology (Weber et al., 2018). The efficient use of an immense amount of data yields enormous benefits when complex systems, such as river and lake systems, are analyzed (Christ et al., 2016). However, water quality prediction with those numerous datasets is still challenging due to multitude of dynamic and interrelated factors, including climate conditions, landuse and seasonal changes. There have been many efforts for developing numerical water quality models such as open-source model of high transferability (Delft3D), Environmental Fluid Dynamics Code (EFDC), and Water Quality Analysis Simulation Program (WASP), but the complexity and uncertainty associated with accurately representing the multitude of processes that influence water quality have been always issued (Beck, 1987; Lindenschmidt, 2006; Xu et al., 2022a,b). Recently, deep learning modeling has become a fruitful approach that has been widely used in water quality modeling. Machine learning models find, or extract features, that is, independent combinations of measurable variables that serve as better predictors or classifiers. These features provide a high-level representation of complex

* Corresponding author. E-mail address: khcho80@korea.ac.kr (K.H. Cho).

https://doi.org/10.1016/j.wroa.2023.100207

Received 13 September 2023; Received in revised form 8 November 2023; Accepted 14 November 2023 Available online 16 November 2023

2589-9147/© 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).







datasets and relationships within them. Such models have been successfully used to estimate various water quality variables including total nitrogen (Sha et al., 2021), total phosphorus (Zhang et al., 2020a), dissolved oxygen (Moghadam et al., 2021), chlorophyll-a (Guo et al., 2022), total organic carbon (Zhu et al., 2019), turbidity (Wan et al., 2022), and colored dissolved organic matter (Niu et al., 2021).

In the original applications of machine learning to predictions of environmental time series such as water quality and metrological data, cells of artificial neural networks (ANNs) having parameters that needed to be determined to provide the best fit used sections of fixed length as inputs. Eventually, it was found that substantially better predictions in time series could be provided with recurrent neural networks (RNNs). A cell of RNNs uses not only current datasets-predictors as inputs but also the output of the cell obtained after its application to the previous input dataset. Thus, previous information is better connected to the present prediction task (Olah, 2015). In applications to time series, RNNs use data sequences of fixed length as the dataset-predictors and can simulate the effect of distant past events on current data in theory. However, in practice, training the RNN, that is, changing its parameters to fit the data, may be stalled because of the unavoidably limited accuracy of the computations. For the same reason, the parameter-search process may vary. These phenomena are known as vanishing gradient and exploding gradient problems. These problems have been alleviated by the introduction of Long Short-Term Memory (LSTM) neural networks (Sherstinsky, 2020). LSTM models are not sensitive to the fact that the time interval between the current observation and the event affecting this observation is unknown and can be large. This capability of LSTM is caused by utilizing the cell state variables that have not existed in other ANN architectures and has allowed selective remembering and forgetting of the sequential information of data (Xu et al., 2022a,b). LSTM is the most cited neural network architecture introduced in the 20th century (Schmidhuber, 2021).

Several studies have been conducted to predict the quality of inland water by applying the LSTM model (e.g., Hu et al., 2019; Lee and Lee, 2018; Liu et al., 2019; Wang et al., 2017). Water quality modeling has expanded to multi-dimensional data, such as multi-spectral imagery, radar images, and GIS-based data (Srivastava et al., 2022). Sagan et al. (2020) successfully applied LSTM model with reflectance spectra data of Landsat-8 and Sentinel-2 to estimate water quality variables such as cyanobacteria, algae, total suspended solid, turbidity, nitrate nitrogen, phosphate phosphorus, and total dissolved solid. In addition, it was found that LSTM performance could be improved by combining LSTM with different deep learning techniques. In particular, the increasing complexity of water-quality-related data sequences has led to the need for feature extraction improvements (Fan et al., 2022; Hao et al., 2022). One of the successful methods of feature extraction is implemented in convolutional neural networks (CNNs) (Wen et al., 2017). The combinations of LSTM and CNNs have been applied in many research fields, demonstrating performance improvement compared with the use of LSTM alone (Kim and Cho, 2019; Vidal and Kristjanpoller, 2020; Zhao et al., 2019). Transfer learning is another method used to build robust deep learning models by extracting data features more flexibly (Abubakr et al., 2022). The weights of the previously developed model were used as reliable initial values for training with similar datasets (Wahlang et al., 2022). The flexibility and robustness of the transfer learning of LSTM have been proven in several studies (Bashar et al., 2020; Jung et al., 2019; Ma et al., 2020a,b). Moreover, attention networks have been applied to improve the feature extraction of deep learning models (Xie et al., 2022). The selective refinement of the learnable parameters-weights in the attention mechanism emphasizes the important features of the data, contributing to the improvement of the LSTM model performance (Kumar et al., 2020; Shen et al., 2020; Zhang et al., 2020a).

This paper reviews applications of the LSTM model for water quality modeling of inland waters. We review the opportunities of LSTM performance enhancement in water quality applications, identify promising avenues for further research, and indicate possible advances in water quality management using LSTM-based techniques.

2. LSTM for water quality modeling

2.1. LSTM theory

To make the LSTM model capable of learning long-term dependencies in a time series, the LSTM cell has a more complex design than that of conventional ANNs or RNNs (Lim and Zohren, 2021; Sezer et al., 2020). Specifically, the LSTM cell has an internal state variable, which, along with the LSTM output from the previous time step, is used as the input for this cell at the current time step. The equations controlling the information flow in the LSTM cell are given below, and complementary Fig. 1(a) and (b) illustrates.

$$\Gamma_f = \sigma \left(W_f \left[h^{\langle t-1 \rangle}, x^{\langle t \rangle} \right] + b_f \right), \tag{1}$$

$$c_f^{(t)} = \Gamma_f \odot C^{(t-1)} \tag{2}$$

$$\Gamma_i = \sigma \big(W_i \big[h^{(t-1)}, x^{(t)} \big] + b_i \big), \tag{3}$$

$$c_c^{(t)} = tanh \big(W_c \big[h^{(t-1)}, x^{(t)} \big] + b_c \big), \tag{4}$$

$$c_i^{(t)} = \Gamma_i \odot c_c^{(t)} \tag{5}$$

$$C^{(t)} = c_f^{(t)} + c_i^{(t)}, \tag{6}$$

$$\Gamma_o = \sigma \big(W_o \big[h^{(t-1)}, x^{(t)} \big] + b_o \big), \tag{7}$$

$$h^{} = \Gamma_o \odot \tanh c^{}$$
(8)

where C indicates the cell state variable, *h* be the hidden state variable (also known as the cell output variable), and *x* be the input dataset. All input and output variables $h^{(t-1)}$, $C^{(t-1)}$, $x^{(t)}$, $C^{(t)}$, and $h^{(t)}$ as well as the intermediate variables $c_f^{(t)}$, $c_c^{(t)}$, and $c_i^{(t)}$ are vectors. The LSTM cell includes operators called 'gates' that modify the information as it moves through the cell. Eqs. (2), (5), and (8) describe the work of the forget, input, and output gates denoting to Γ_f , Γ_i , and Γ_o respectively. The arguments of the signoidal functions include the vectors of biases b_f , b_i , or b_o , and the products of matrices of weights W_f , W_i , and W_o and the vector $[h^{(t-1)}, x^{(t)}]$, which is the concatenation of the vector of hidden state values $h^{(t-1)}$ and input vector $x^{(t)}$. The symbol \odot denotes the Hadamard product of the two vectors when elements corresponding to the same row are multiplied. At each step, an LSTM cell receives $x^{(t)}$ as inputs and produces $C^{(t)}$ and $h^{(t)}$ as outputs.

2.2. Water quality simulation of LSTM and process-driven numerical models

A wide range of process-driven modeling tools are available which can simulate spatial and temporal evolution of water quality parameters in surface waters. These include Hydrological Simulation Program FORTRAN (HSPF) (Bicknell et al., 1997), WASP (Wool et al., 2020), Integrated Catchment (INCA) (Whitehead et al., 2016), enhanced stream water quality model (QUAL2E) (Brown and Barnwell, 1987), one-dimensional river and stream water quality model (QUAL2K), one-dimensional river model (MIKE11) (Havnø et al., 1995), EFDC (Hamrick, 1992) and Delft3D (Roelvink and Van Banning, 1995) among others. These are time-proven models which have been used in hundreds of water quality simulation studies worldwide (Tong et al., 2022). To simulate fate and transport of conservative substances, these models incorporate modeling of advection and dispersion phenomena. For non-conservative pollutants, both chemical and biological



Fig. 1. Architecture of LSTM cell (a) Information flow according to the LSTM model Eqs. (1)-(6). Numbers of equations are in parentheses. Explanation of symbols is in the text following the equations. (b) The commonly used representation of the LSTM cell explicitly showing the nonlinear transformations within the LSTM cell.

transformations are taken into consideration. While models like HSPF, INCA, and Soil and Water Assessment Tool (SWAT) are suitable for conducting 1D simulations of surface water quality constituents at the catchment scale, more detailed and comprehensive simulation outcomes can be achieved through the use of numerical models such as WASP, Delft3D, and EFDC (Bouraoui and Grizzetti, 2014; Keller et al., 2023; Zhou et al., 2023).

One advantage of numerical water quality models over LSTM is that they can provide detailed 3-dimensional fate and transport of pollutants in geometrically complex water bodies (Cho et al., 2020). This detailed simulation can provide profound insights into the behavior of pollutants in aquatic environments. However, these high-resolution simulations come at the cost of computation resources. Furthermore, these numerical models leverage mass conservation equations which forces these models to follow laws of nature during simulation (Abbaspour et al., 2015). The ability to simulate detailed processes in complex environments while following physical laws increases representative capacity of these numerical models (Clark et al., 2015). However, the effectiveness of numerical models because of their high representation of underlying processes depends on discretization step used to determine the grid. While a smaller grid size can simulate detailed processes however it exponentially increases the computation demand (Leng et al., 2022). On the other hand, larger grid size can omit important details thereby compromising model's efficacy. Despite these advantages, one limiting factor in the application of these numerical models is their requirement for detailed input information. This information is often required in the form of boundary or initial conditions which are often difficult to measure. Another limitation of these models is that they cannot incorporate a wide variety of input data. On the other hand, LSTM based deep learning models can process numerical, image, or text information simultaneously (Ward et al., 2020). This flexibility of LSTM has the potential to build more representative water quality models by incorporating wide range of factors affecting the target variable (Brunner et al., 2021).

Another advantage of numerical models over LSTM based deep learning models is the interpretation of parameters of calibrated models (Tsai et al., 2021). The parameters of a well calibrated numerical water quality model reveal important insights about the physical world (Feng et al., 2022). For example, parameters K12C, K12T and KNIT in WASP are related to nitrification while CCHL, HNRP and KMNG correspond to nutrient update by algae (Wool et al., 2020). Similarly, parameters KT and BM of EFDC model, describe effect of temperature and basal metabolism on cyanobacteria (Hamrick, 1992). However, a similar interpretation of an LSTM model with high prediction performance is not possible. Research has been carried out to interpret either the parameters of trained LSTM or intermediate outputs from a trained LSTM model. Kratzert et al. (2019a,b) showed correspondence between hydrological system (snow-water equivalent and soil moisture) and LSTM behavior. A similar interpretation of a trained LSTM for water quality parameters can also be performed in future.

2.3. Performance of LSTM compared with other machine learning models

LSTM models have demonstrated performance in water quality prediction than other machine learning models by showing higher correlation coefficient and lower error. AlDahoul et al. (2021) conducted experiments to evaluate LSTM's performance compared to other models. The first trial predicted sediment concentration for the next day using 10 years of data. Four models-ElasticNet LR, Multilayer Perceptron Neural Network, Extreme Gradient Boosting, and LSTM-were tested, with LSTM consistently outperforming others. LSTM also competed with ElasticNet LR in predicting monthly sediment levels despite a smaller dataset. In the third experiment, LSTM excelled in forecasting sediment concentrations seven days ahead. Overall, LSTM's predictive power exceeded ElasticNet LR's across scenarios. Dong et al. (2021) showed LSTM's superiority over traditional neural networks in predicting windspeed data collected every five minutes over a year. LSTM's success stemmed from capturing nonlinear features in data sequences, unlike traditional neural networks (Lagesse et al., 2020; Li et al., 2017). Shewalkar (2019) highlighted LSTM's advantages in speech recognition over various methods, citing its computational efficiency, high accuracy, and robustness. In essence, the collective evidence highlights LSTM's dominant performance across various domains and challenges, underscoring its potential as a powerful tool in predictive modeling, particularly for time-series and sequence-dependent data.

2.4. LSTM enhanced by preprocessing

Environmental data have static and dynamic variables with different kinds and different time scales. Thus, Data preprocessing is crucial for enhancing the performance of LSTM models due to noise reduction, feature scaling, and improving convergence. Recent studies have shown that preprocessing the original data prior to training yields improved LSTM results (Sha et al., 2021; Sun and Huang, 2020). In a study conducted by Sha et al. (2021), they employed the Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) method to separate noise and periodic components from data. These extracted components were used as inputs for the LSTM model, resulting in improved model performance. However, the application of CEEM-DAN had differing effects on non-periodic (total nitrogen, TN) and strong-periodic (dissolved oxygen, DO) water quality parameters, with TN improvements being significantly greater than those for DO.

Dong and Zhang (2021) introduced a two-stage decomposition technique involving CEEMDAN and variational mode decomposition (VMD) to analyze polycyclic aromatic hydrocarbon concentration time series. The LSTM was then applied to explore the characteristics of each subsequence, leading to enhanced prediction performance. Notably, the additional application of VMD further improved predictive capabilities. Despite CEEMDAN's efficiency with complex and variable data, room for improvement in predictive performance remains due to inadequately extracting high-frequency features from signals. Hence, the CEEM-DAN-VMD two-stage decomposition approach maximizes information from the PAH time series and significantly enhances model prediction performance (Li et al., 2021). Wang et al. (2023) introduced a short-term water quality prediction model designed to address the nonlinear, unstable, and random nature of water quality parameters. This model combines VMD to enhance LSTM performance. Initially, VMD decomposes water quality data into stable components, reducing data instability and improving predictability. Each component is then used in the LSTM model for prediction.

Nourani and Behfar (2021) introduced two LSTM variants, namely SLSTM and WLSTM, for estimating suspended sediment load (SSL) at gauging stations in the Missouri and Upper Mississippi areas. SLSTM captures time-series seasonality of the elements. WLSTM employs wavelet transform to decompose signals into distinct patterns for input. Both models effectively identify SSL's long-term properties, enhancing performance (Foufoula-Georgiou et al., 1994).

The Synchrosqueezed Wavelet Transform (SWT, Daubechies et al., 2011) enhances time-frequency resolution of non-stationary signals by reorganizing the continuous wavelet transform's time-frequency map along the frequency axis. SWT also exhibits robustness against noise (Herrera et al., 2014). Song et al. (2021) integrated LSTM with SWT to denoise water quality data and enhance model performance. Using 365 weekly dissolved oxygen (DO) values from two stations in the Haihe River Basin, they showcased improved outcomes through the noise-resistant capabilities of SWT and LSTM's nonlinear mapping. A limitation was the insufficient data to fully explore all of SWT's advantages.

3. LSTM combination with CNN

LSTM networks have been pivotal in modeling time-series water quality due to their ability to simulate long-range dependences, but have limitations when applied standalone to water quality prediction. In particular, using LSTMs on multiple-dimensional data of water quality can result in inefficient resource usage. Barzegar et al. (2020) used the low performance of LSTM standalone model to predict DO and Chl-a in the Small Prespa Lake in Greece. This is because combining LSTM and CNN can provide significant improvement of model accuracies by leveraging complementary data.

3.1. LSTM-CNN theory

A hybrid model combining CNN and LSTM improves feature learning from complex inputs. A CNN is a feedforward neural network that uses deep convolutional operations to enhance feature extraction (Kim and Cho, 2019; Yan et al., 2021). A typical CNN structure includes a convolutional layer, pooling layer, fully connected layer, and an output layer. The convolutional layer performs computations with learnable weights and biases in the convolutional filters for input transformation and feature extraction (Eq. (9)). The pooling layer conducts downsampling of convolutional features. This can reduce the training cost of the model owing to the lower number of parameters, and the max-pooling layer extracts the maximum values from the convolutional features (Eq.10). Finally, the fully connected layer integrates the feature information by flattening the local feature maps into a one-dimensional feature vector, and then delineates it to the output layer (Eq. (11)).

$$x^{l} = f_{a} \left(\sum w^{l} \odot x^{l-1} + b^{l} \right)$$
(9)

$$x_{ij}^{l} = Max_{k=0,..n,s=0,..n} \left(x_{(i+k)(j+s)}^{l} \right)$$
(10)

$$y = f_a \left(w x_{flat}^l + b \right) \tag{11}$$

where x^{l} is the convolutional operation output in the lth layer, x^{l-1} is the convolutional output in the *l*-1th layer, f_{a} is the activation function, w^{l} is the learnable weight in the lth layer, b^{l} is the bias in the *l*th layer, Max indicates the max-pooling operator, x_{ij}^{l} is the maximum output in elements *I* and *j* in the lth layer, x_{flat}^{l} is the flattened output of the lth convolutional layer, and *y* is the output.

In Fig. 2(b), the nodes of the fully connected neurons from the CNN model are fed to the LSTM cell, as shown in Fig. 1. The vector of the CNN outputs $y_{conv}^{(l)}$ plays the role of the vector $x^{(l)}$ in Fig. 1. The way of combining the CNN and LSTM models can vary depending on the research purpose, input dimensionality, and feature extraction priority. One representative hybrid network is the LSTM-CNN model (Fig. 2(b)). In this network structure, LSTM extracts time-series information of the input data, and then the periodic features are used as the input of the consecutive CNN model training for classification or regression tasks. LSTM cell output $h^{(l)}$ serves as the input for the CNN.

3.2. LSTM-CNN application for water quality prediction

The CNN-LSTM experiment predicting water quality was conducted by Baek et al. (2020) in the Nakdong River basin, simulating TN, total phosphorus, and total organic carbon levels. Using radar images, temperature, evaporation rate, and other inputs, the CNN model predicted water levels. These water levels, along with prior water quality data, were fed into the LSTM model, yielding CNN-LSTM having R^2 values above 0.86 for training and over 0.79 for validation. In the Presa basin, Barzegar et al. (2020) employed a CNN-LSTM model to predict DO and chlorophyll a concentrations. The CNN model extracted periodic features from pH, conductivity, and other variables, which were then amplified by the LSTM. Training accuracy for DO and Chl-a reached correlation coefficients of 0.96 and 0.92, with validation at 0.97 and 0.87. The hybrid CNN-LSTM also outperformed other models. Sha et al. (2021) validated robust DO and TN predictions in the Xin'anjiang River using CNN-LSTM. Employing 12-day input data and employing complementary ensemble empirical mode decomposition with adaptive noise (CEEMDAN), the CNN-LSTM model demonstrated improved performance compared to standalone CNN and LSTM models. The preprocessing further enhanced predictions, improving Nash-Sutcliffe efficiency values for both DO and TN. In summary, these studies collectively showcase CNN-LSTM's capability in short-term water quality forecasting and highlight the impact of preprocessing on its performance.

Jang et al. (2021) utilized the LSTM-CNN model to forecast antibiotic-resistance gene (ARG) occurrences at Gwangalli Beach, Republic of Korea. Cumulative rainfall, tides, salinity, temperature, wind parameters, and humidity formed the input time-series data. LSTM initially extracted essential features, followed by CNN further refining these features to enhance ARG prediction. The LSTM-CNN model outperformed the sole LSTM, achieving R^2 values between 0.20 and 0.67 for training and 0.15 to 0.55 for validation. The authors emphasized the LSTM-CNN's potential for predicting contaminants and water quality variables in aquatic settings. Furthermore, Yao et al. (2023) presented a data-driven deep learning model for runoff prediction, combining CNN-LSTM to improve prediction accuracy and versatility. The model



Fig. 2. Combination of the LSTM and CNN networks to hybrid models (a) CNN-LSTM and (b) LSTM-CNN.

Conv layer

uses meteorological, hydrological, and runoff data, reducing dimensionality with the maximum information coefficient. It employs CNN for long time series feature extraction, LSTM for long-term runoff prediction. Also, Liu et al. (2023) showed approach combining a two-stage feature selection of a hybrid deep learning model including CNN, LSTM, effectively captures the complex relationships in multivariate time series data from wastewater treatment plants (WWTPs). A two-stage feature selection process optimizes the feature subset to enhance prediction accuracy of predicting effluent total nitrogen in WWTPs. models for forecasting both water quality and contaminants in aquatic environments. Thus, the ensemble LSTM with CNN model offered a valuable approach for improving prediction accuracy with additional feature extraction of the data. Herein, CNN-LSTM specialized to deal with spatial data while LSTM-CNN particularized the time-series information.

These studies underscore the efficacy of coupling CNN and LSTM

4. LSTM with attention mechanism

4.1. LSTM-attention theory

Yet another method that can enhance feature extraction in a data series is including the attention mechanism in the LSTM model. The attention mechanism can improve the selection of input sequences and encode information in long-term memory (Li et al., 2019). Although LSTM addresses the long-term dependency problem, a vanishing gradient problem of network training may arise. This makes it difficult to capture distant past data, which may be valuable predictors. This implies that the performance of LSTM may not be satisfactory if important information is in the distant past from the time when the predictions are made. To avoid this issue, the LSTM model can focus on the relevant information using the attention mechanism rather than simply learning it in chronological order (Zheng et al., 2021b). Traditional LSTM, however, uses a simple concatenation of $h^{< t-1>}$ and $x^{< t>}$ as $[h^{< t-1>}, x^{< t>}]$ (Fig. 1); the LSTM with attention applies attention weights to $x^{<t>}$ before concatenation, and these weights are different for different gates. Attention weights are being determined during the training process. Through the back-propagation process, the attention weights are being optimized along with the weights of the LSTM.

4.2. Attention-LSTM application for water quality prediction

Recent studies have employed attention-LSTM in water quality modeling. Zhu et al. (2021) used a fusion model incorporating ResNet, BiLSTM, and multihead attention to predict DO concentrations in Lake Taihu. The attention mechanism extracted features from diverse environmental data, yielding superior results compared to LSTM without attention. Cao et al. (2021) applied attention-LSTM to simulate DO concentrations in ponds, achieving an RMSE of 0.380 mg/L. Cao et al. (2021) also explored attention's impact on the simpler GRU model, finding the attention-GRU outperformed both conventional GRU and attention-LSTM models. Jang et al. (2021) integrated input attention (IA) with LSTM for predicting ARG occurrence at a beach. IA-LSTM improved over conventional LSTM, though LSTM-CNN excelled in certain scenarios. IA-LSTM outperformed for multi-ARG simulations, while LSTM-CNN proved superior for single ARG predictions. This suggests attention mechanisms enhance LSTM's learning in intricate contexts.

Overall, the attention-LSTM model outperformed the conventional LSTM model in simulating water quality. The advantage of attention-LSTM models is that they can identify the relevance of information, which is not possible with the conventional LSTM model. However, attention-enhanced LSTM models are not always superior to other model combinations; therefore, the dataset-specific model structure must be carefully selected.

5. LSTM combination with transfer learning (TL)

Most water quality data are automatically measured by sensors at multiple monitoring stations in water-environment (Zhang et al., 2019). Numerous reasons (e.g., data entry errors, non-response, technical issues, and natural cases) may cause substantial missing values in raw water quality datasets (Ma et al., 2020a). These missing data significantly degrade water quality prediction using LSTM. Therefore, Transfer learning could be an alternative way to resolve limitations associated with missing water quality data.

5.1. LSTM-TL theory

Transfer learning (TL) is a methodology that transfers learned knowledge from one problem (source) to another but similar problem (target). This technique has been applied when a model is trained with insufficient or missing datasets. This can be achieved by transferring knowledge from the source domain (i.e., complete data) to the target domain (i.e., incomplete data); for example, TL can transfer knowledge from the complete data obtained at adjacent monitoring stations to missing data in the target domain (Chen et al., 2021). Transfer learning can be categorized into data pattern transfer (e.g., trend and statistical characteristics), model transfer (e.g., model structure and parameters), and task transfer (Pan and Yang, 2009). In LSTM applications, data pattern and model transfers have been used to overcome the missing-data problem (Chen et al., 2021; Zhou 2020). Data pattern transfer selects the source domain that has the greatest statistical similarity with the target domain, and both datasets are mixed and used to train the LSTM model. Model transfer means that the optimized structure (multiple output and number of hidden layers) and parameters (learning rate, weight vector, and bias vector) from the source domain can be transferred to the LSTM model in the target domain (Zhou, 2020).

5.2. LSTM-TL application for water quality predictions

Zhou (2020) employed data pattern and model transfers to counteract missing data effects. Comparing general LSTM and TL-LSTM, they assessed various missing-data rates and forecast scenarios. TL-LSTM's effectiveness was demonstrated in both data missing rates and forecast lead-time situations. For instance, at a 10-day lead and 90 % data missing rate, TL-LSTM improved RMSE and NSE by 24.7 % and 23.3 % respectively. Chen et al. (2021) used TL-LSTM with TrAdaBoost to impute missing DO concentration data, achieving 15–25 % better imputation performance than other methods (Fig. 3). Similarly, Tian et al. (2019) optimized LSTM for Chl-a concentration prediction using TL, outperforming parameter norm penalties and dropout methods over a 3-month period. Thus, these studies highlight the potential of transfer learning to enhance water quality modeling and prediction accuracy under varying conditions.

6. Influence of data transformation and static inputs on the model performance

Data transformation plays an important role in determining the performance of LSTM for water quality prediction. The selection of data transformation methods can substantially influence how good the LSTM captures patterns, handles temporal dependencies of water quality. Abbas et al. (2022) observed the influence of data transformations on the model performance. The results of LSTM with the logarithmic input transformation were closer to the observations than those of the min-max transformation, yielding an NSE of 0.57 (Fig. 4). A negative Percent Bias value was obtained after logarithmic transformation. This indicated that the simulated *Escherichia coli* from the logarithmic transformation was underestimated, whereas the result of the min-max transformation was overestimated. This behavior may be attributed to the higher sensitivity of min-max scaling to outliers (Chuang et al., 2010).

Static inputs can have a significant influence on the performance of LSTM models in water quality prediction. These static inputs do not change over time and can provide additional information to the LSTM model. Incorporating static inputs into LSTM models can improve their ability to capture complex relationships and increase prediction accuracy. Efforts to integrate static attributes into LSTM have been explored. Lees et al. (2021a) introduced an entity-aware LSTM (EA-LSTM) for hydrological pattern learning based on static catchment features, yet their study indicated a performance decline compared to basic LSTM. Abbas et al. (2023) incorporated sub-basin characteristics as static data alongside continuous input data for LSTM. To address static data's lack of temporal information, they initialized LSTM's hidden and cell states with static values, assuming this strategy would provide predictive insights.

To investigate the importance of static input data, they compared the performance of the model with an LSTM model without static data. The



Fig. 3. Proposed large-scale consecutive missing-data imputation transfer algorithm based on TrAdaBoost-LSTM (From Chen et al., 2021).

results also indicated that, without using static variables, the performance of LSTM was worse. Another example, Frame et al. (2021) utilized LSTM with both hydrological static and dynamic data to improve the accuracy of predicting stream flow. They noted that altering the routing mechanism could preserve temporal details from the land surface processes, while adjustments to the evapotranspiration choices might maintain information related to mass bias. These imply that the use of static physical features of sub-basins to initialize hidden and cell states provides important contextual information to the LSTM. This information regarding the physical features of a sub-basin improves the LSTM prediction performance at the validation site. The simultaneous learning of between time series and static physical features leverages the ability of LSTM training and evade the unrealistic parametrization of applying traditional deterministic modeling (Kratzert et al., 2019a,b). In addition, using the static input feature of LSTM attribute to direct application for the cases that have similar regional and physical information (Drost et al., 2021).

7. Summary: LSTM-based approaches for water quality simulation

We reviewed various LSTM-based models used in water quality analysis and their key features, along with potential challenges and concerns associated with each model. Table 1 presents concise summary of each model. CEEDAN-LSTM preprocesses the data to distinguish and remove noise from periodic elements, then applies this processed data as input to an LSTM model. Care should be taken when interpreting how CEEMDAN components and LSTM states interact, particularly in the context of water quality analysis. VMD-LSTM effectively manages multiscale and non-linear temporal patterns in water quality data. There is a potential risk of overfitting due to incomplete regularization and validation. CEEDAN-VMD-LSTM employs a dual-stage decomposition method to optimize information attributes. This model demands careful focus on hyperparameter adjustment, data preparation, and model interpretation. Sequenced-LSTM employs multiple LSTMs to capture autoregressive elements from the original water quality data. There are possible challenges related to vanishing and exploding gradients should be considered. Wavelet-LSTM breaks down signals into various distinct patterns through the application of wavelet functions. It is crucial to carefully select the wavelet function and effectively manage boundary effects and sensitivity to noise. CNN-LSTM adopts convolutional layers to extract features from the data. This approach may introduce increased model complexities, necessitating meticulous hyperparameter tuning. LSTM-CNN utilizes CNN as a post-processing step within the model. This may lead to heightened model complexities and requires thorough attention to hyperparameter tuning. Attention-LSTM concentrates on specific segments of input sequences during processing via attention mechanisms. There is a potential for overfitting and heightened computational demands that should be considered. TL-LSTM helps leverage information from one domain to improve performance in another but challenging when dealing with mismatched domains.

Each model has its unique strengths and challenges, and careful consideration of these features and cautions is essential when choosing the appropriate model for a specific water quality simulation.



Fig. 4. Influence of data transformation on the model performance (Abbas et al., 2022) in terms of correlation, standard deviation; the closer to the reference (red star), the better model performance.

8. Outlook: challenges and opportunities

8.1. Integrating multi-sources, extension, and generalization

Water quality is influenced by various sources including land use, climate, and agricultural activities. Incorporating data from multisources (e.g., static covariates or diverse time-varying variables) into LSTM models is one of major challenges. Therefore, it is very crucial to apply robust feature selection and dimensionality reduction methods to handle multi-sources data.

The LSTM approach can generate predictions using only forward data information, whereas a bi-directional LSTM (Bi-LSTM) method considers both forward and backward neural network information to make predictions. The Bi-LSTM effectively captures past and future information independently at specific time steps through its dual LSTM. Thus, by incorporating both positive and negative time series data rules, the extension of LSTM model, Bi-LSTM, may achieve higher prediction accuracy compared to LSTM (Tan et al., 2023).

LSTM models trained with specific water quality station might struggle to generalize to another stations with distinct characteristics even in same watershed. Here, Transferability and accurate prediction of LSTM across multiple stations in same or different watersheds remain challenging. Effectively incorporating transfer learning by pre-training on similar water quality station could be one of solutions to the resolve the issue. Leveraging domain adaptation techniques can help the LSTM generalize successfully (Deng et al., 2022)

Table 1

Features and a cautionary note of LSTM-based models.

Model	Features	A cautionary note
CEEDAN- LSTM	A preprocessing process for extracting to separate the separate the noise from periodic components of the environmental data and used or insure of the LTM model	Interpreting interactions between CEEMDAN components and LSTM states (e. g., water quality)
VMD-LSTM	A preprocessing process to effectively handle and model multi-scale and non-linear temporal patterns within water quality-related data	Overfitting problem caused by incomplete regularization and validation
CEEMDAN-	A two-stage decomposition	High level of care and attention
VMD-LSTM	technology maximizes the information characteristics	to detail in hyperparameter tuning, data preparation, and model interpretation
Sequenced- LSTM	Usage of multiple LSTMs to extracts the autoregressive component of the original water quality data	Vanishing and Exploding Gradients
Wavelet- LSTM	A method to decompose a signal into multiple and special pattern	The choice of wavelet function and handling boundary effects and noise sensitivity.
CNN-LSTM	Usage of CNN for feature extraction from the data	Higher complexities and hyperparameter tuning
LSTM-CNN	Usage of CNN as a post- processing process.	Higher complexities and hyperparameter tuning
Attention- LSTM	Focusing on specific parts of input sequences during processing	Overfitting and computation demands
TL-LSTM	Helping leverage information from one domain to improve performance in another.	Mismatched domains

8.2. Model interpretability and explainability

Interest in explaining the mechanism of deep learning models is increasing (Floridi, 2021). This is more the case when making high-stake decisions based on the results of LSTM-based modeling (Lipton, 2018). Therefore, it is not only important to make good predictions using LSTM models but also to explain them. The methods used to explain the decision-making process of LSTM can be categorized into two types: 1) model-agnostic interpretation and 2) model-dependent interpretation (Molnar et al., 2020a). Model-agnostic interpretation methods consider the LSTM as a black box and explain its behavior by analyzing its output against perturbed inputs (Molnar et al., 2020b). However, these methods have also received criticism due to inconsistencies and instability in their results (Rudin, 2019). On the other hand, model-dependent interpretation methods rely on the internal state of LSTM, including the weights, cell state, and hidden state of LSTM. These methods employ post-hoc approaches, such as integrating the gradients or using simple linear regression (SLR) on the internal elements of LSTM (i.e., cell and hidden states) to explain the decision-making process in LSTM (Lees et al., 2021b). The use of SLR to explain the behavior of deep learning models is common in natural language processing (Hewitt and Liang, 2019). Recently, Lees et al. (2021b) conducted a study in which the cell state of an LSTM was interpreted with the help of SLR. The objective of this study was to determine whether LSTM can learn a hydrologically realistic phenomenon during training. They found that, during the training of LSTM to model the rainfall-runoff process, the internal elements of LSTM learn an intermediate phenomenon (e.g., changes in water storage in the soil). The study showed that the cell state of LSTM after the application of SLR can represent soil moisture. The use of SLR to interpret the internal variables of LSTM can be extended to water quality modeling. LSTM has been successfully used to model micropollutants and harmful algal blooms in surface waters (Yun et al., 2021; Zheng et al., 2021a). However, these models only provide target prediction at a specific site and do not indicate whether the LSTM has learned a representative related-phenomenon. The use of SLR along with LSTM, as shown by Lees et al. (2021b), can be adopted for this purpose.

Another approach used for making LSTM results more tenable is to force the model to follow physical laws by encoding these natural laws inside the LSTM. One such example is the mass conservation LSTM (MC-LSTM), in which the mass conservation principle is encoded in the equations of the LSTM (Hoedt et al., 2021). In the MC-LSTM, the mass of the target input inside the cell state is conserved over time. This is performed by combining the gating mechanism of LSTM and the redistribution of input in the cell state of LSTM. The authors argue that MC-LSTM can be extended to conserve variables such as energy, momentum, and count. Because energy and mass conservation laws are fundamental parts of water quality models (Neitsch et al., 2011), the use of MC-LSTM can make physically consistent and realistic water quality predictions.

8.3. Handling availability and data quality

A major challenge in the applicability of LSTM for water quality modeling is the scarcity of observation data. Monitoring and analyzing of water quality parameters is time consuming and costly (Baek et al., 2021). However, the development of data-driven models with high prediction performance requires a large amount of data (Goodfellow et al., 2016). There have been attempts to overcome this shortcoming by incorporating domain knowledge into the LSTM along with input data. This mitigates the potential breaches of physical laws that might arise due to limited input data (Zhang et al., 2020). The domain knowledge is given to the model in the form of constraints and boosts on the loss function which makes the solution within a feasible and realistic solution space (Jia et al., 2018; Karpatne et al., 2017; Raissi, 2018; Raissi et al., 2019).

In water quality modeling, even when large amounts of data are available, they are usually not monitored at a uniform time step, which is an essential requirement for time-series modeling using LSTM Chollet, 2018). Usually, water quality is monitored at irregular time steps or during certain seasons or special time intervals, such as during rainfall events (Jang et al., 2021). The LSTM enhances with ordinary differential equation (ODE-LSTM) has been specifically designed to model irregularly sampled time-series data (Rubanova et al., 2019). In the ODE-LSTM, the states of LSTM are considered as continuous time representations and modeled by solving the ODE at every time step using a numerical ODE solver (Lechner and Hasani, 2020). This means that, instead of using standard LSTM equations (Eqs. (1)-((6) to calculate its states, ODE-LSTM solves the initial value problem between two observations and calculates the LSTM state (Eq. (6)) (Rubanova et al., 2019). This feature of ODE-LSTM makes it a natural tool for modeling irregularly sampled water quality data.

8.4. Efficient optimization

Optimizing LSTM models helps in finding the best combination of hyperparameters that can exploit the unique capabilities of LSTMs to their fullest extent, leading to more accurate predictions, reduced training time, and ultimately better performance on various tasks involving sequential data. In particular, employing machine learning techniques for optimizing LSTM models enhances the efficiency, effectiveness, and performance of these models, making them more suitable for real-world applications involving sequential data. Dheda et al. (2022) introduced Genetic Algorithm (GA) optimization with LSTM to enhance DO prediction performance of Burnnett River and Baffle River in Australia. In GA-LSTM model, throughout the process of optimization, the genetic operators were utilized to explore the potential solutions within the search space. This led to the formation of a population consisting of conceivable solutions, represented as chromosomes encoded using binary bits. These chromosomes conveyed information about the optimal hyperparameter values, specifically indicating the quantity of LSTM units and the size of the time window. The LSTM models that underwent optimization using the GA demonstrated superior performance compared to their initial versions. This improvement in model performance was achieved by fine-tuning hyperparameters through the GA optimization process. The most significant outcome of GA optimization was the reduction in both the overall computational time required and the count of parameters that could be trained in the models.

8.5. Fusion with statistical and process-based model

The combination of LSTM and statistical modeling explores the synergistic integration of deep learning and statistical techniques in inland water environment. This approach aims to leverage LSTM's ability to capture complex temporal dependencies in data alongside the strengths of statistical modeling, which provides interpretability, robustness, and probabilistic insights. This fusion helps accomplish time series prediction, anomaly detection, and uncertainty quantification for water environment analysis. Thus, unique perspective on the fusion of LSTM and statistical inference opens up the potential for improvement of data analysis and predictive modeling (Fang et al., 2022).

Theory-Guided machine learning is the other promising approach by combining process-based and LSTM models. Frame et al. (2021) proposed LSTM postprocessors providing significant benefit to enhance the performance of United States National Water Model (NWM). Employing LSTM for post-processing offers the potential to enhance the interpretability. Their findings revealed a strong correlation between the representation of hydrologic characteristics in the LSTM model and NWM. This suggests that the functions learned by the LSTM, which map inputs to streamflow, exhibit substantial similarity. Furthermore, even though articulating the intricate LSTM behavior concisely through compact formulas like partial differential equations might be challenging due to the numerous trained model parameters, LSTM model can recognize their structural resemblances to process-based models such as the NWM.

In conclusion, the future research of LSTM in the water environment depends on tackling these bottlenecks through interdisciplinary and innovative collaboration. Tackling these challenges will lead to more accurate, robust, and practical LSTM-based water quality prediction models, contributing to effective water resource management and environmental protection.

9. Conclusion

The major findings from our review are as follows:

- LSTM models demonstrate robust performance for water quality prediction compared with other machine learning models. They can extract the dynamic features of water quality, store the most valuable time-frequency information, and achieve accurate predictions.
- Combining LSTM and CNN improves predictions by both letting CNN to process LSTM output and letting LSTM to process CNN output.
- Being an efficient prediction tool, LSTM can be further enhanced by combining it with attention mechanism and transfer learning approach.
- Better interpretation of results and efficient use of relevant non-time series-type information present both challenges and opportunities for the development of the LSTM-based data analysis and prediction tools in water quality area.
- Using the watershed static information and the data transformation method can substantially enhance the LSTM performance.

Declaration of Competing Interest

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

Data availability

Data will be made available on request.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) [No. 2022R1A2C2006172] and this work also partially supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2021R1A6A1A03039572)

References

- Abbas, A., Boithias, L., Pachepsky, Y., Kim, K., Chun, J.A., Cho, K.H., 2022. AI4Water v1.0: an open-source python package for modeling hydrological time series using data-driven methods. Geosci. Model. Dev. 15 (7), 3021–3039. https://doi.org/ 10.5194/gmd-15-3021-2022.
- Abbas, A., Park, M., Baek, S.S., Cho, K.H., 2023. Deep learning-based algorithms for longterm prediction of chlorophyll-a in catchment streams. J. Hydrol. 626, 130240. https://doi.org/10.1016/j.jhydrol.2023.130240.
- Abbaspour, K.C., Rouholahnejad, E., Vaghefi, S., Srinivasan, R., Yang, H., Kløve, B., 2015. A continental-scale hydrology and water quality model for Europe: calibration and uncertainty of a high-resolution large-scale SWAT model. J. Hydrol. 524, 733–752. https://doi.org/10.1016/j.jhydrol.2015.03.027.
- Abubakr, M., Akoush, B., Khalil, A., Hassan, M.A., 2022. Unleashing deep neural network full potential for solar radiation forecasting in a new geographic location with historical data scarcity: a transfer learning approach. Eur. Phys. J. Plus 137 (4), 474. https://doi.org/10.1140/epip/s13360-022-02666-y.
- AlDahoul, N., Essam, Y., Kumar, P., Ahmed, A.N., Sherif, M., Sefelnasr, A., Elshafie, A., 2021. Suspended sediment load prediction using long short-term memory neural network. Sci. Rep. 11 (1), 7826. https://doi.org/10.1038/s41598-021-87415-4.
- Baek, S.-S., Choi, Y., Jeon, J., Pyo, J., Park, J., Cho, K.H., 2021. Replacing the internal standard to estimate micropollutants using deep and machine learning. Water Res. 188, 116535 https://doi.org/10.1016/j.watres.2020.116535.
- Baek, S.-S., Pyo, J., Chun, J.A., 2020. Prediction of water level and water quality using a CNN-LSTM combined deep learning approach. Water 12 (12), 3399. https://doi.org/ 10.3390/w12123399.
- Barzegar, R., Aalami, M.T., Adamowski, J., 2020. Short-term water quality variable prediction using a hybrid CNN–LSTM deep learning model. Stoch. Environ. Res. Risk Assess. 34 (2), 415–433. https://doi.org/10.1007/s00477-020-01776-2.
- Bashar, M.A., Nayak, R., Suzor, N., 2020. Regularising LSTM classifier by transfer learning for detecting misogynistic tweets with small training set. Knowl. Inf. Syst. 62 (10), 4029–4054. https://doi.org/10.1007/s10115-020-01481-0.
- Beck, M.B., 1987. Water quality modeling: a review of the analysis of uncertainty. Water Resour. Res. 23 (8), 1393–1442. https://doi.org/10.1029/WR023i008p01393.
- Bicknell, B.R., Imhoff, J.C., Kittle Jr., J.L., Donigian Jr., A.S., Johanson, R.C, 1997. Hydrological Simulation Program—FORTRAN user's Manual for Version 11. US Environmental Protection Agency, Athens, GA. Environmental Protection Agency Report No. EPA/600/R-97/080.
- Bouraoui, F., Grizzetti, B., 2014. Modelling mitigation options to reduce diffuse nitrogen water pollution from agriculture. Sci. Total Environ. 468–469, 1267–1277. https:// doi.org/10.1016/j.scitotenv.2013.07.066.
- Brown, L.C., Barnwell, T.O., 1987. The Enhanced Stream Water Quality Models QUAL2E and QUAL2E-UNCAS: Documentation and User Manual. US Environmental Protection Agency. Office of Research and Development. Environmental Research Laboratory, Athens, p. 189.
- Brunner, M.I., Slater, L., Tallaksen, L.M., Clark, M., 2021. Challenges in modeling and predicting floods and droughts: a review. WIREs Water 8 (3), e1520. https://doi. org/10.1002/wat2.1520.
- Cao, X., Ren, N., Tian, G., Fan, Y., Duan, Q., 2021. A three-dimensional prediction method of dissolved oxygen in pond culture based on attention-GRU-GBRT. Comput. Electron. Agric. 181, 105955 https://doi.org/10.1016/j.compag.2020.105955.
- Chen, Z., Xu, H., Jiang, P., Yu, S., Lin, G., Bychkov, I., Hmelnov, A., Ruzhnikov, G., Zhu, N., Liu, Z., 2021. A transfer learning-based LSTM strategy for imputing largescale consecutive missing data and its application in a water quality prediction system. J. Hydrol. 602, 126573 https://doi.org/10.1016/j.jhydrol.2021.126573.
- Cho, K.H., Pachepsky, Y., Ligaray, M., Kwon, Y., Kim, K.H., 2020. Data assimilation in surface water quality modeling: a review. Water Res. 186, 116307 https://doi.org/ 10.1016/j.watres.2020.116307.

Chollet, F., 2018. Deep Learning With Python. Manning Publications Co.

- Christ, M., Kempa-Liehr, A.W. & Feindt, M. (2016). Distributed and parallel time series feature extraction for industrial big data applications. arXiv preprint arXiv.161 0.07717 https://doi.org/10.48550/arXiv.1610.07717.
- Chuang, C.-C., Wang, C.-M., Li, C.-W., 2010. Weighted Linear Regression for Symbolic Interval-Values Data With Outliers. IEEE, pp. 2238–2242. https://doi.org/10.1109/ ICIEA.2010.5515157.
- Clark, M.P., Fan, Y., Lawrence, D.M., Adam, J.C., Bolster, D., Gochis, D.J., Hooper, R.P., Kumar, M., Leung, L.R., Mackay, D.S., Maxwell, R.M., Shen, C., Swenson, S.C.,

Zeng, X., 2015. Improving the representation of hydrologic processes in Earth system models. Water Resour. Res. 51 (8), 5929–5956. https://doi.org/10.1002/2015WR017096.

- Daubechies, I., Lu, J., Wu, H.-T., 2011. Synchrosqueezed wavelet transforms: an empirical mode decomposition-like tool. Appl. Comput. Harmon Anal. 30 (2), 243–261. https://doi.org/10.1016/j.acha.2010.08.002.
- Deng, R., Panl, L., Clavel, C., 2022. Domain adaptation for stance detection towards unseen target on social media. In: 2022 10th International Conference on Affective Computing and Intelligent Interaction (ACII). IEEE, pp. 1–8. https://doi.org/ 10.1109/ACII55700.2022.9953818.
- Dheda, D., Cheng, L., Abu-Mahfouz, A.M., 2022. Long short term memory water quality predictive model discrepancy mitigation through genetic algorithm optimisation and ensemble modeling. IEEE Access 10, 24638–24658. https://doi.org/10.1109/ ACCESS.2022.3152818 f.
- Dong, L., Zhang, J., 2021. Predicting polycyclic aromatic hydrocarbons in surface water by a multiscale feature extraction-based deep learning approach. Sci. Total Environ. 799, 149509 https://doi.org/10.1016/j.scitotenv.2021.149509.
- Dong, Y., Zhang, Y., Liu, F., Cheng, X., 2021. Reservoir production prediction model based on a stacked LSTM network and transfer learning. ACS Omega 6 (50), 34700–34711. https://doi.org/10.1021/acsomega.1c05132.
- Drost, S., Mudersbach, C., 2021. The impact of land cover data on rainfall-runoff prediction using an entity-aware-LSTM. In: EGU General Assembly Conference Abstracts, pp. EGU21–E1136. https://doi.org/10.5194/egusphere-egu21-1136.
- Fan, C., Li, Y., Yi, L., Xiao, L., Qu, X., Ai, Z., 2022. Multi-objective LSTM ensemble model for household short-term load forecasting. Memet. Comput. 14 (1), 115–132. https://doi.org/10.1007/s12293-022-00355-y.
- Feng, D., Liu, J., Lawson, K., Shen, C., 2022. Differentiable, learnable, regionalized process-based models with multiphysical outputs can approach state-of-the-art hydrologic prediction accuracy. Water Resour. Res. 58 (10), e2022WR032404 https://doi.org/10.1029/2022WR032404.
- Fang, K., Kifer, D., Lawson, K., Feng, D., Shen, C., 2022. The data synergy effects of timeseries deep learning models in hydrology. Water Resour. Res. 58 (4), e2021WR029583 https://doi.org/10.1029/2021WR029583.
- Floridi, L., 2021. The European Legislation on AI: a brief analysis of its philosophical approach. Philos. Technol. 34 (2), 215–222. https://doi.org/10.1007/978-3-031-09846-8 1.
- Foufoula-Georgiou, E., Kumar, P., Chui, C., 1994. Wavelets in Geophysics. Academic Press.
- Frame, J.M., Kratzert, F., Raney, A., Rahman, M., Salas, F.R., Nearing, G.S., 2021. Postprocessing the national water model with long short-term memory networks for streamflow predictions and model diagnostics. JAWRA Journal of the American Water Resources Association 57 (6), 885–905. https://doi.org/10.1111/1752-1688.12964.
- Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press.
- Guo, H., Tian, S., Huang, J.J., Zhu, X., Wang, B., Zhang, Z., 2022. Performance of deep learning in mapping water quality of Lake Simcoe with long-term Landsat archive. ISPRS J. Photogramm. Remote Sens. 183, 451–469. https://doi.org/10.1016/j. isprsiors.2021.11.023.
- Hao, W., Sun, X., Wang, C., Chen, H., Huang, L., 2022. A hybrid EMD-LSTM model for non-stationary wave prediction in offshore China. Ocean Eng. 246, 110566 https:// doi.org/10.1016/j.oceaneng.2022.110566.

Hamrick, J.M. (1992). A three-dimensional environmental fluid dynamics computer code: theoretical and computational aspects. https://doi.org/10.21220/V5TT6C.

- Havnø, K., Madsen, M., Dørge, J., 1995. MIKE 11-a generalized river modelling package. Computer Models of Watershed Hydrology. Water Resources Publications, Colorado, USA, pp. 733–782.
- Herrera, R.H., Han, J., van der Baan, M., 2014. Applications of the synchrosqueezing transform in seismic time-frequency analysis. Geophysics 79 (3), V55–V64. https:// doi.org/10.1190/geo2013-0204.

Hewitt, J. & Liang, P. (2019). Designing and interpreting probes with control tasks. arXiv preprint arXiv.1909.03368 https://doi.org/10.48550/arXiv.1909.03368.
Hoedt, P.-J., Kratzert, F., Klotz, D., Halmich, C., Holzleitner, M., Nearing, G.S.,

- Hoedt, P.-J., Kratzert, F., Klotz, D., Halmich, C., Holzleitner, M., Nearing, G.S., Hochreiter, S., Klambauer, G., 2021. MC-LSTM: Mass-Conserving LSTM. PMLR, pp. 4275–4286.
- Hu, Z., Zhang, Y., Zhao, Y., Xie, M., Zhong, J., Tu, Z., Liu, J., 2019. A water quality prediction method based on the deep LSTM network considering correlation in smart mariculture. Sensors 19 (6), 1420. https://doi.org/10.3390/s19061420.
- Jang, J., Abbas, A., Kim, M., Shin, J., Kim, Y.M., Cho, K.H., 2021. Prediction of antibiotic-resistance genes occurrence at a recreational beach with deep learning models. Water Res. 196, 117001 https://doi.org/10.1016/j.watres.2021.117001.
- Jia, X., Karpatne, A., Willard, J., Steinbach, M., Read, J., Hanson, P.C., Dugan, H.A. & Kumar, V. (2018). Physics guided recurrent neural networks for modeling dynamical systems: application to monitoring water temperature and quality in lakes. arXiv preprint arXiv.1810.02880 https://doi.org/10.48550/arXiv.1810.02880.
- Jung, S., Park, J., Lee, S., 2019. Polyphonic Sound Event Detection Using Convolutional Bidirectional LSTM and Synthetic Data-Based Transfer Learning. IEEE, pp. 885–889. https://doi.org/10.1109/ICASSP.2019.8682909.
- Karpatne, A., Watkins, W., Read, J. & Kumar, V. (2017). Physics-guided neural networks (PGNN): an application in lake temperature modeling. arXiv preprint arXiv.1710.11 431 https://doi.org/10.48550/arXiv.1710.11431.
- Keller, A.A., Garner, K., Rao, N., Knipping, E., Thomas, J., 2023. Hydrological models for climate-based assessments at the watershed scale: a critical review of existing hydrologic and water quality models. Sci. Total Environ. 867, 161209 https://doi. org/10.1016/j.scitotenv.2022.161209.

Kim, T.-Y., Cho, S.-B., 2019. Predicting residential energy consumption using CNN-LSTM neural networks. Energy 182, 72–81. https://doi.org/10.1016/j. energy.2019.05.230.

Kumar, A., Narapareddy, V.T., Srikanth, V.A., Malapati, A., Neti, L.B.M., 2020. Sarcasm detection using multi-head attention based bidirectional LSTM. IEEE Access 8, 6388–6397. https://doi.org/10.1109/ACCESS.2019.2963630.

Kratzert, F., Herrnegger, M., Klotz, D., Hochreiter, S., Klambauer, G., 2019a. Explainable AI: Interpreting, Explaining and Visualizing Deep Learning. Springer, pp. 347–362.

Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., Nearing, G., 2019b. Benchmarking a catchment-aware long short-term memory network (LSTM) for large-scale hydrological modeling. Hydrol. Earth Syst. Sci. Discuss. 2019, 1–32.

Lagesse, B., Wang, S., Larson, T.V., Kim, A.A., 2020. Predicting PM2.5 in well-mixed indoor air for a large office building using regression and artificial neural network models. Environ. Sci. Technol. 54 (23), 15320–15328. https://doi.org/10.1021/acs. est.0c02549.

Lechner, M. & Hasani, R. (2020). Learning long-term dependencies in irregularlysampled time series. arXiv preprint arXiv.2006.04418 https://doi.org/10.4855 0/arXiv.2006.04418.

Lee, S., Lee, D., 2018. Improved prediction of harmful algal blooms in four major South Korea's rivers using deep learning models. Int. J. Environ. Res. Public Health 15 (7), 1322. https://doi.org/10.3390/ijerph15071322.

Lees, T., Buechel, M., Anderson, B., Slater, L., Reece, S., Coxon, G., Dadson, S.J., 2021a. Benchmarking data-driven rainfall-runoff models in Great Britain: a comparison of long short-term memory (LSTM)-based models with four lumped conceptual models. Hydrol. Earth Syst. Sci. 25 (10), 5517–5534. https://doi.org/10.5194/hess-25-5517-2021.

Lees, T., Reece, S., Kratzert, F., Klotz, D., Gauch, M., De Bruijn, J., Kumar Sahu, R., Greve, P., Slater, L., Dadson, S., 2021b. Hydrological concept formation inside long short-term memory (LSTM) networks. Hydrol. Earth Syst. Sci. Discuss. 1–37. https:// doi.org/10.5194/hess-26-3079-2022.

Leng, L., Xu, C., Jia, H., Jia, Q., 2022. Incorporating receiving waters responses into the framework of spatial optimization of LID-BMPs in plain river network region. Water Res. 224, 119036 https://doi.org/10.1016/j.watres.2022.119036.

Li, C., Wang, P., Wang, S., Hou, Y., Li, W., 2017. Skeleton-Based Action Recognition Using LSTM and CNN 585–590. https://doi.org/10.1109/ICMEW.2017.8026287.

Li, H., Jin, F., Sun, S., Li, Y., 2021. A new secondary decomposition ensemble learning approach for carbon price forecasting. Knowledge-Based Syst. 214, 106686 https:// doi.org/10.1016/j.knosys.2020.106686.

Li, Y., Zhu, Z., Kong, D., Han, H., Zhao, Y., 2019. EA-LSTM: evolutionary attention-based LSTM for time series prediction. Knowledge-Based Syst. 181, 104785 https://doi. org/10.1016/j.knosys.2019.05.028.

Lim, B., Zohren, S., 2021. Time-series forecasting with deep learning: a survey. Philos. Trans. R. Soc. A 379 (2194), 20200209. https://doi.org/10.1098/rsta.2020.0209.

Lindenschmidt, K.E., 2006. The effect of complexity on parameter sensitivity and model uncertainty in river water quality modelling. Ecol. Model. 190 (1–2), 72–86. https:// doi.org/10.1016/j.ecolmodel.2005.04.016.

Lipton, Z.C., 2018. The mythos of model interpretability: in machine learning, the concept of interpretability is both important and slippery. Queue 16 (3), 31–57. https://doi.org/10.1145/3236386.3241340.

Liu, W., Liu, T., Liu, Z., Luo, H., Pei, H., 2023. A novel deep learning ensemble model based on two-stage feature selection and intelligent optimization for water quality prediction. Environ. Res. 224, 115560 https://doi.org/10.1016/j. envires.2023115560

Liu, P., Wang, J., Sangaiah, A.K., Xie, Y., Yin, X., 2019. Analysis and prediction of water quality using LSTM deep neural networks in IoT environment. Sustainability 11 (7), 2058. https://doi.org/10.3390/su11072058.

Ma, J., Cheng, J.C., Jiang, F., Chen, W., Wang, M., Zhai, C., 2020a. A bi-directional missing data imputation scheme based on LSTM and transfer learning for building energy data. Energy Build. 216, 109941 https://doi.org/10.1016/j. enbuild.2020.109941.

Ma, J., Cheng, J.C., Ding, Y., Lin, C., Jiang, F., Wang, M., Zhai, C., 2020b. Transfer learning for long-interval consecutive missing values imputation without external features in air pollution time series. Adv. Eng. Inform. 44, 101092 https://doi.org/ 10.1016/j.aei.2020.101092.

Moghadam, S.V., Sharafati, A., Feizi, H., Marjaie, S.M.S., Asadollah, S.B.H.S., Motta, D., 2021. An efficient strategy for predicting river dissolved oxygen concentration: application of deep recurrent neural network model. Environ. Monit. Assess 193 (12), 1–18. https://doi.org/10.1007/s10661-021-09586-x.

Molnar, C., Casalicchio, G., Bischl, B., 2020a. Interpretable Machine Learning—A Brief History, State-of-the-Art and Challenges. Springer, pp. 417–431.

Molnar, C., König, G., Herbinger, J., Freiesleben, T., Dandl, S., Scholbeck, C.A., Casalicchio, G., Grosse-Wentrup, M. & Bischl, B. (2020b). Pitfalls to avoid when interpreting machine learning models. arXiv preprint arXiv:2007.04131.Neitsch, S.L., Arnold, J.G., Kiniry, J.R., Williams, J.R., 2011. Soil and Water Assessment

Tool Theoretical Documentation Version 2009. Texas Water Resources Institute. Niu, C., Tan, K., Jia, X., Wang, X., 2021. Deep learning based regression for optically

inactive inland water quality parameter estimation using airborne hyperspectral imagery. Environ. Pollut. 286, 117534 https://doi.org/10.1016/j. envpol.2021.117534.

Nourani, V., Behfar, N., 2021. Multi-station runoff-sediment modeling using seasonal LSTM models. J. Hydrol. 601, 126672 https://doi.org/10.1016/j. ihydrol.2021.126672.

Olah, C. (2015). Understanding lstm networks.

Pan, S.J., Yang, Q., 2009. A survey on transfer learning. IEEE Trans. Knowl. Data Eng. 22 (10), 1345–1359. https://doi.org/10.1109/TKDE.2009.191. Raissi, M., 2018. Deep hidden physics models: deep learning of nonlinear partial differential equations. J. Mach. Learn. Res. 19 (1), 932–955. http://jmlr.org/pap ers/v19/18-046.html.

Raissi, M., Perdikaris, P., Karniadakis, G.E., 2019. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. J. Comput. Phys. 378, 686–707. https://doi. org/10.1016/i.icn.2018.10.045.

Roelvink, J., Van Banning, G., 1995. Design and development of DELFT3D and application to coastal morphodynamics. Oceanogr. Lit. Rev. 11 (42), 925. Rubanova, Y., Chen, R.T., Duvenaud, D.K., 2019. Latent ordinary differential equations

for irregularly-sampled time series. Adv. Neural Inf. Process. Syst. 32, 1–11. Rudin, C., 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat. Mach. Intell. 1 (5), 206–215. https://doi.org/10.1038/s42256-019-0048-x.

Sagan, V., Peterson, K.T., Maimaitijiang, M., Sidike, P., Sloan, J., Greeling, B.A., Adams, C., 2020. Monitoring inland water quality using remote sensing: potential and limitations of spectral indices, bio-optical simulations, machine learning, and cloud computing. Earth-Science Rev. 205, 103187 https://doi.org/10.1016/j. earscirev.2020.103187.

Schmidhuber, J., 2021. AI Blog. IDSIA, Switzerland. Retrieved 2022-04-30.

Sezer, O.B., Gudelek, M.U., Ozbayoglu, A.M., 2020. Financial time series forecasting with deep learning: a systematic literature review: 2005–2019. Appl. Soft Comput. 90, 106181 https://doi.org/10.1016/j.asoc.2020.106181.

Sha, J., Li, X., Zhang, M., Wang, Z.-L., 2021. Comparison of forecasting models for realtime monitoring of water quality parameters based on hybrid deep learning neural networks. Water 13 (11), 1547. https://doi.org/10.3390/w13111547.

Shen, Z., Zhang, Q., Han, K., Huang, D.-S., 2020. A deep learning model for RNA-protein binding preference prediction based on hierarchical LSTM and attention network. IEEE/ACM Trans. Comput. Biol. Bioinform. 753–762. https://doi.org/10.1109/ TCBB.2020.3007544.

Sherstinsky, A., 2020. Fundamentals of recurrent neural network (RNN) and long shortterm memory (LSTM) network. Physica D 404, 132306. https://doi.org/10.1016/j. physd.2019.132306.

- Shewalkar, A, 2019. Performance evaluation of deep neural networks applied to speech recognition: RNN, LSTM and GRU. J. Artif. Intell. Soft Comput. Res. 9 (4), 235–245. https://doi.org/10.2478/jaiscr-2019-0006.
- Song, C., Yao, L., Hua, C., Ni, Q., 2021. A novel hybrid model for water quality prediction based on synchrosqueezed wavelet transform technique and improved long shortterm memory. J. Hydrol. 603, 126879 https://doi.org/10.1016/j. ihydrol.2021.126879.
- Srivastava, A., Singhal, A., Jha, P.K., 2022. Ecological Significance of River Ecosystems. Elsevier, pp. 105–132.

Sun, W., Huang, C., 2020. A novel carbon price prediction model combines the secondary decomposition algorithm and the long short-term memory network. Energy 207, 118294. https://doi.org/10.1016/j.energy.2020.118294.

Tan, R., Wang, Z., Wu, T., Wu, J., 2023. A data-driven model for water quality prediction in Tai Lake, China, using secondary modal decomposition with multidimensional external features. J. Hydrol. 47, 101435 https://doi.org/10.1016/j. eirh.2023.101435.

Tian, W., Liao, Z., Wang, X., 2019. Transfer learning for neural network model in chlorophyll-a dynamics prediction. Environ. Sci. Pollut. Res. 26 (29), 29857–29871. https://doi.org/10.1007/s11356-019-06156-0.

Tong, X., Mohapatra, S., Zhang, J., Tran, N.H., You, L., He, Y., Gin, K.Y.-H., 2022. Source, fate, transport and modelling of selected emerging contaminants in the aquatic environment: current status and future perspectives. Water Res. 217, 118418 https://doi.org/10.1016/j.watres.2022.118418.

Tsai, W.-P., Feng, D., Pan, M., Beck, H., Lawson, K., Yang, Y., Liu, J., Shen, C., 2021. From calibration to parameter learning: harnessing the scaling effects of big data in geoscientific modeling. Nat Commun 12 (1), 5988. https://doi.org/10.1038/ s41467-021-26107-z

Vidal, A., Kristjanpoller, W., 2020. Gold volatility prediction using a CNN-LSTM approach. Expert Syst. Appl. 157, 113481 https://doi.org/10.1016/j. eswa.2020.113481.

Wahlang, I., Maji, A.K., Saha, G., Chakrabarti, P., Jasinski, M., Leonowicz, Z., Jasinska, E., 2022. Brain magnetic resonance imaging classification using deep learning architectures with gender and age. Sensors 22 (5), 1766. https://doi.org/ 10.3390/s22051766.

Wan, S., Yeh, M.-L., Ma, H.-L., Chou, T.-Y., 2022. The robust study of deep learning recursive neural network for predicting of turbidity of water. Water 14 (5), 761. https://doi.org/10.3390/w14050761.

Wang, Y., Zhou, J., Chen, K., Wang, Y., Liu, L., 2017. Water Quality Prediction Method Based on LSTM Neural Network 1–5. https://doi.org/10.1109/ISKE.2017.8258814.

Wang, Z., Wang, Q., Wu, T., 2023. A novel hybrid model for water quality prediction based on VMD and IGOA optimized for LSTM. Front. Environ. Sci. Eng. 17 (7), 88. https://doi.org/10.1007/s11783-023-1688-y.

Ward, N.D., Megonigal, J.P., Bond-Lamberty, B., Bailey, V.L., Butman, D., Canuel, E.A., Diefenderfer, H., Ganju, N.K., Goñi, M.A., Graham, E.B., Hopkinson, C.S., Khangaonkar, T., Langley, J.A., McDowell, N.G., Myers-Pigg, A.N., Neumann, R.B., Osburn, C.L., Price, R.M., Rowland, J., Sengupta, A., Simard, M., Thornton, P.E., Tzortziou, M., Vargas, R., Weisenhorn, P.B., Windham-Myers, L., 2020. Representing the function and sensitivity of coastal interfaces in Earth system models. Nat. Commun. 11 (1), 2458. https://doi.org/10.1038/s41467-020-16236-2.

Weber, L.J., Muste, M., Bradley, A.A., Amado, A.A., Demir, I., Drake, C.W., Krajewski, W. F., Loeser, T.J., Politano, M.S., Shea, B.R., 2018. The Iowa watersheds project: Iowa's prototype for engaging communities and professionals in watershed hazard J. Pyo et al.

mitigation. Int. J. River Basin Manag. 16 (3), 315–328. https://doi.org/10.1080/ 15715124.2017.1387127.

Wen, L., Li, X., Gao, L., Zhang, Y., 2017. A new convolutional neural network-based datadriven fault diagnosis method. IEEE Trans. Ind. Electron. 65 (7), 5990–5998. https://doi.org/10.1109/TIE.2017.2774777.

- Whitehead, P.G., Leckie, H., Rankinen, K., Butterfield, D., Futter, M., Bussi, G., 2016. An INCA model for pathogens in rivers and catchments: model structure, sensitivity analysis and application to the River Thames catchment, UK. Sci. Total Environ. 572, 1601–1610. https://doi.org/10.1016/j.scitotenv.2016.01.128.
- Wool, T., Ambrose Jr., R.B., Martin, J.L., Comer, A, 2020. WASP 8: the next generation in the 50-year evolution of USEPA's water quality model. Water 12 (5), 1398. https:// doi.org/10.3390/w12051398.
- Xie, Y., Feng, D., Shen, X., Liu, Y., Zhu, J., Hussain, T., Baik, S.W., 2022. Clustering feature constraint multiscale attention network for shadow extraction from remote sensing images. IEEE Trans. Geosci. Remote Sens. 60, 1–14. https://doi.org/ 10.1109/TGRS.2022.3151901.
- Xu, S., He, G., Fang, H., Bai, S., Wu, X., 2022a. Parameter uncertainty and sensitivity analysis of the three Gorges Reservoir and Xiangxi River EFDC model. J. Hydrol. 610, 127881 https://doi.org/10.1016/j.jhydrol.2022.127881.
- Xu, Y., Hu, C., Wu, Q., Jian, S., Li, Z., Chen, Y., Zhang, G., Zhang, Z., Wang, S., 2022b. Research on particle swarm optimization in LSTM neural networks for rainfall-runoff simulation. J. Hydrol. 608, 127553 https://doi.org/10.1016/j.jhydrol.2022.127553.
- Yan, R., Liao, J., Yang, J., Sun, W., Nong, M., Li, F., 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.
- Yao, Z., Wang, Z., Wang, D., Wu, J., Chen, L., 2023. An ensemble CNN-LSTM and GRU adaptive weighting model based improved sparrow search algorithm for predicting runoff using historical meteorological and runoff data as input. J. Hydrol. 625, 129977 https://doi.org/10.1016/j.jhydrol.2023.129977.
- Yun, D., Abbas, A., Jeon, J., Ligaray, M., Baek, S.-S., Cho, K.H., 2021. Developing a deep learning model for the simulation of micro-pollutants in a watershed. J. Clean. Prod. 300, 126858 https://doi.org/10.1016/j.jclepro.2021.126858.

- Zhang, H., Zhang, Q., Shao, S., Niu, T., Yang, X., 2020a. Attention-based LSTM network for rotatory machine remaining useful life prediction. IEEE Access 8, 132188–132199. https://doi.org/10.1109/ACCESS.2020.3010066.
- Zhang, R., Liu, Y., Sun, H., 2020b. Physics-informed multi-LSTM networks for metamodeling of nonlinear structures. Comput. Methods Appl. Mech. Eng. 369, 113226 https://doi.org/10.1016/j.cma.2020.113226.
- Zhang, Y.F., Thorburn, P.J., Xiang, W., Fitch, P., 2019. SSIM—A deep learning approach for recovering missing time series sensor data. IEEE Internet Things J. 6 (4), 6618–6628. https://doi.org/10.1109/JIOT.2019.2909038.
- Zhao, J., Mao, X., Chen, L., 2019. Speech emotion recognition using deep 1D & 2D CNN LSTM networks. Biomed. Signal Process. Control 47, 312–323. https://doi.org/ 10.1016/j.bspc.2018.08.035.
- Zheng, L., Wang, H., Liu, C., Zhang, S., Ding, A., Xie, E., Li, J., Wang, S., 2021a. Prediction of harmful algal blooms in large water bodies using the combined EFDC and LSTM models. J. Environ. Manag. 295, 113060 https://doi.org/10.1016/j. jenvman.2021.113060.
- Zheng, W., Zhao, P., Huang, K., Chen, G., 2021b. Understanding the property of long term memory for the LSTM with attention mechanism. In: Proceedings of the 30th ACM International Conference on Information & Knowledge Management, pp. 2708–2717. https://doi.org/10.1145/3459637.3482399.
- Zhou, Y., 2020. Real-time probabilistic forecasting of river water quality under data missing situation: deep learning plus post-processing techniques. J. Hydrol. 589, 125164.
- Zhou, L., Wu, F., Meng, Y., Byrne, P., Ghomshei, M., Abbaspour, K.C., 2023. Modeling transport and fate of heavy metals at the watershed scale: state-of-the-art and future directions. Sci. Total Environ. 878, 163087 https://doi.org/10.1016/j. scitotenv.2023.163087.
- Zhu, L., Zhang, C., Zhang, C., Zhang, Z., Nie, X., Zhou, X., Liu, W., Wang, X., 2019. Forming a new small sample deep learning model to predict total organic carbon content by combining unsupervised learning with semisupervised learning. Appl. Soft Comput. 83, 105596 https://doi.org/10.1016/j.asoc.2019.105596.
- Zhu, N., Ji, X., Tan, J., Jiang, Y., Guo, Y. 2021. Prediction of dissolved oxygen concentration in aquatic systems based on transfer learning. Comput. Electron. Agric. 180, 105888 https://doi.org/10.1016/j.compag.2020.105888.