Application of Attention mechanism combined with Long Short-Term Memory for forecasting Dissolved Oxygen in Ganga River

Neha Pant[1], Durga Toshniwal[1], Bhola Ram Gurjar[2]

[1]Department of Computer Science and Engineering,
Indian Institute of Technology Roorkee, India
{npant, durga.toshniwal}@cs.iitr.ac.in
[2]Department of Civil Engineering,
Indian Institute of Technology Roorkee, India
bhola.gurjar@ce.iitr.ac.in

Abstract. Accurate forecasting of water quality parameters is a significant part of the process of water resource management. In this paper we demonstrate the applicability of Long Short-Term Memory (LSTM) combined with attention mechanism for the long-term forecasting (after 24 hours) of Dissolved Oxygen content at various stations of Ganga River flowing through the state of Uttar Pradesh, India. In the given model, the hidden states of the LSTM units are passed to the attention layer. The attention layer then gives different weights to the hidden states based on their relevance. The performance of the models is evaluated using root mean square error, mean absolute error and coefficient of determination. The experimental results indicate that combining attention mechanism with LSTM significantly improves the forecasted values of Dissolved Oxygen when compared with state-of-the-art models like Recurrent Neural Network, LSTM, and bidirectional LSTM. The demonstrated model is particularly useful during the availability of only univariate datasets.

Keywords: Dissolved Oxygen, water quality forecasting, Long Short-Term Memory, Attention.

1 Introduction

Water is an essential and irreplaceable resource supporting human, animal, and plant life on earth through its consumption in various forms. Because of rapid urbanization in the recent decades, the number of industrial setups and towns near the rivers has increased substantially in India, causing water quality to deteriorate day by day. This water quality deterioration has called for the efficient management of water resources. Accurate forecasting is an essential step in improving the management of water resources.

Dissolved Oxygen (DO) is one of the many important parameters used to determine the standard of water in rivers and other water bodies. Dissolved Oxygen is the balanced amount of oxygen dissolved in the water bodies left after oxygen-producing processes like absorption from the atmosphere and photosynthesis and oxygen-consuming processes like aerobic respiration and chemical oxidation. Water must contain sufficient DO for the aquatic life to survive. However, continuous anthropogenic activities near the water bodies have led to the depletion of Dissolved Oxygen, thereby resulting in the deterioration of water quality standards.

Currently, many departments managing water resources worldwide have set up sensors to monitor the quality of water at different monitoring stations. However, they cannot be used to forecast the quality of water. However, at the same time, the enormous data collected by these stations can be used to build certain data-based models to predict future values of specific water quality parameters. The forecasts obtained from the models can be used further to control the degree of pollution in the water bodies by encouraging policymakers and government officials to take appropriate actions to manage water resources properly.

In this paper, we have demonstrated the applicability of the attention mechanism combined with LSTM for long-term forecasting of DO content in River Ganga flowing through the state of Uttar Pradesh in India. We also compare the developed model with other state-of-the-art models like Recurrent Neural Network, LSTM, and bidirectional LSTM. The presented method is particularly beneficial when only univariate datasets are available.

2 Related Work

Several studies have used different variations of Artificial Neural Networks for forecasting of different water quality parameters. The employment of artificial neural network (ANN) to determine the amount of biological oxygen demand (BOD) and DO of Gomti river in India has been suggested by Singh et al. in [1]. The study concluded that phosphate, nitrate nitrogen, ammoniacal nitrogen and Chemical Oxygen Demand were the main factors in determining DO. Ruben G. B. et al. [2] recommends the utilization of MLP model with Levenberg-Marquardt algorithm for learning for the forecast the level of COD in Xuxi River, China. I. S. Yeon et al. [3] proposes the use of Levenberg-Marquardt Neural Network (LMNN), Modular Neural Network (MDNN) and Adaptive Neuro-Fuzzy Inference System (ANFIS) models for forecasting DO. Rankovic et al. [4] have conducted a study making use of a feed forward neural network (FFNN) to estimate the strength of DO in the Gruza Reservoir of Serbia. S. Emamgholizadeh et al. [5] suggests the use of various kinds of neural networks like Multi-Layer Perceptron (MLP), Radial basis network (RBF) and ANFIS for water quality forecasting. It is found in the study that pH and temperature have the highest contribution in improving efficiency of the model whereas the nitrates, phosphates and chlorides have been useless in determining the level of DO. Sarkar A. et al. [6] have proposed a study making the use of ANN with feed forward error back propagation to determine the level of DO in Yamuna River in the downstream of Mathura city in India. The authors suggested that better forecasting is possible if higher frequency data is available for training purpose.

A successful demonstration of wavelet neural network (WNN) and artificial neural network models with different combinations of input parameters with different time lags was used to predict DO, temperature, and salinity by Alizadeh et al [7]. A study comparing linear and non-linear models to forecast the DO levels in the river Danube has been conducted by Csábrági, A. et al. [8]. For this, the authors have employed four different models namely, multi-linear regression, multi-layer perceptron neural network (MLPNN), Radial Basis Function Neural Network, and General Regression Neural Network (GRNN).

Several advanced neural networks have also been used in several studies to find out the complex relationship between the input and output parameters. Zou Q. et al. [9] suggested the use of multi-time scale bidirectional LSTM for prediction of water quality in Beilun River. The authors made use of both water quality and meteorological features in the prediction process. Li et al., 2018 [10] proposed the use of combination of Sparse Auto-encoder and LSTM for predicting DO in a shrimp pond by making use of multivariate dataset containing both water quality and meteorological parameters.

Attention based mechanism was originally used for Natural Language Processing [11]. Nevertheless, the attention mechanism has recently been used for forecasting in several areas like wind power [12], flood [13], electrical loads [14], etc. Liu et al., 2019 [15] makes use of recurrent neural network combined with attention on short and long-term prediction of DO in a pond in Zhejiang Institute of Freshwater Fisheries, Zhejiang province, China. The study makes use of DO, soil parameters and meteorological parameters for short term and long term DO forecasting. The earlier works have made use of historic data consisting of multiple water quality metrics and other external factors like meteorological features. There is a need for methods to be able to handle the non-linear complex relationship between the input and the output features while considering minimal historic data for forecasting DO levels.

3 Data and Methods

3.1 Study area and water quality data

River Ganga is considered one of the most sacred rivers to the Hindus. The river flows for a length of around 2525 KM, flowing through the states of Uttarakhand, Uttar Pradesh, Bihar, and West Bengal in India. The dataset contains values of Dissolved Oxygen for a total of 5 monitoring stations of River Ganga flowing through the Uttar Pradesh region of India. The data is collected from monitoring stations at Ghatiya Ghat bridge (Farrukhabad), Manimau bridge in Kannauj, Bridge at Bithoor, Bridge near Fatehpur (Asni Village) and Bridge in Varanasi. The studied locations are marked on Google maps in Fig. 1. The water quality data is collected from the Uttar Pradesh Pollution Control Board, Lucknow, Uttar Pradesh.

The dataset contains records from 1 April 2017 to 30 April 2021. The dataset available has a sampling rate of one observation per hour for the years 2017 to 2018 and a sampling rate of one observation per 15 minutes for the years 2019 to 2021. In order to ensure consistency of the sampling rate, we retrieved the records with a sample rate of one observation per hour for all the years. Table 1 illustrates the total number of records in the dataset for each of the monitoring stations. The dataset is divided in such a way that 80 percent of the data is used for training and 20 percent is used for testing.

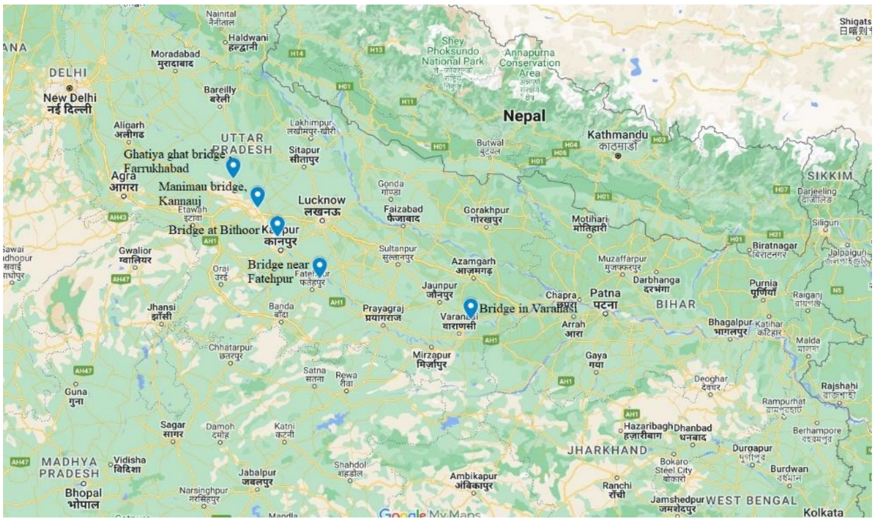


Fig. 1. Study area marked on Google maps

3.2 Data Preprocessing

The dataset contains some random missing values. These missing values are imputed using linear interpolation. The data is further normalized using min-max normalization, thus mapping the values in the range [0,1]. The following is the equation used for normalizing the dataset:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

Where, x' is the normalized value, x is the original value, x_{min} is the minimum value and x_{max} is the maximum value of the variable.

Table 1. Number of records in the dataset for the monitoring stations

Station	Number of records
Ghatiya Ghat bridge, Farrukhabad	31970
Manimau bridge, Kannauj	32298
Bridge at Bithoor	33526
Bridge near Fatehpur	35094
Bridge in Varanasi	34938

3.3 Recurrent Neural Network

Recurrent Neural Networks (RNN) are an advancement over simple neural networks making use of sequential data as input. While simple neural network considers input and output to be independent, RNNs takes into account the previous input sequence to determine the current output. The problem with a simple RNN is that it suffers from vanishing and exploding gradients.

3.4 Long Short-Term Memory

To solve the problem of vanishing gradient and gradient explosion, Hochreiter et al. [16] proposed LSTMs. The LSTMs are a special kind of RNNs, with the ability to learn long-term dependencies by storing information related to different time periods using cell states. The cell state at a specific LSTM unit describes the information that has been considered relevant till that particular timestamp. The LSTMs regulate the flow of information to the cell states by making use of forget, input and output gates. The information passed to the gates is a function of hidden states passed from the previous LSTM unit at the previous timestamps and information at the current timestamp.

3.5 Bidirectional Long Short-Term Memory

Bidirectional LSTM is an advancement over the basic LSTM model. The input is passed to the two LSTMs: first through the forward layer and then through the backward layer. In the forward layer, the LSTM is applied on the input in the forward direction and in the backward layer, another LSTM is applied to the input in the backward direction. After learning the sequence in both the directions, merging operation is performed on the two models.

3.6 Long Short-Term Memory with Attention

Originally, attention mechanism was intended for use in Natural Language processing [11] but was quickly adopted in other disciplines as well [12-14]. The basic idea in the attention mechanism is to consider only the most relevant information and to reduce the impact of less important information from further processing.

The LSTM units extract the long-term dependencies. In temporal attention model, the hidden states of the LSTM units are passed to the attention layer. The attention layer assigns different weights to the various hidden states based on their significance for forecasting dissolved oxygen, thus further enhancing the performance of the model. The features thus generated are then further passed to the fully connected layer which finally generates the forecasts. Figure 2 shows the block diagram for our proposed framework.

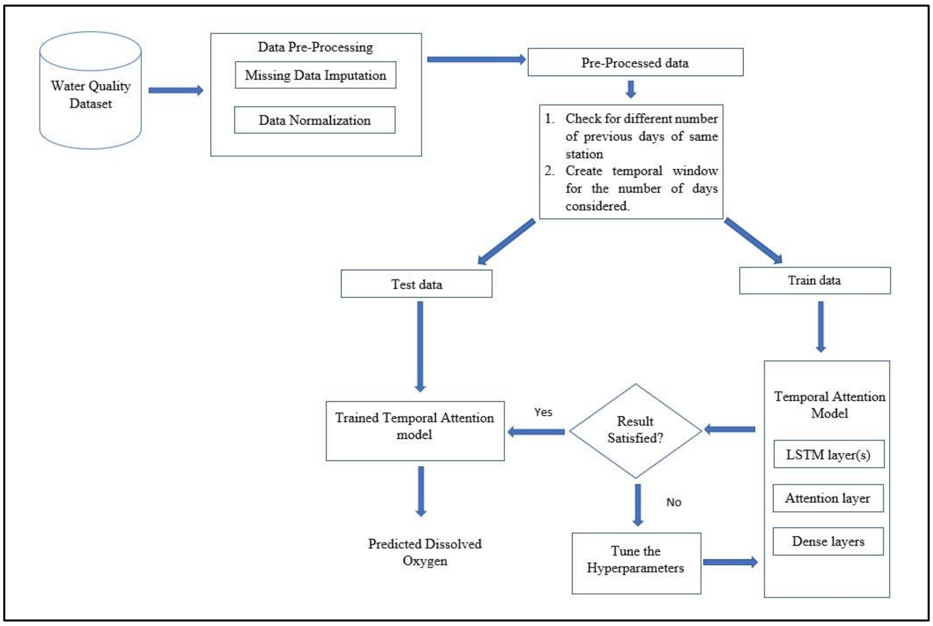


Fig. 2. Block diagram for the proposed framework

4 Results and Discussions

The work has been implemented in Python 3.8 using TensorFlow 2.3.1 and Keras 2.4.3. All the experiments are conducted in Windows 10 operating system with 16 GB RAM and Intel(R) Core (TM) i7-10750H CPU at 2.60GHz. Adam optimizer is used to accelerate the gradient descent algorithm to minimize the mean squared error. The activation function used is the tanh function for all hidden layers.

The performance of the models is checked on learning rates [0.01, 0.001, 0.0001] and on the batch sizes of [64, 128, 256]. The models perform best with the learning rate of 0.0001 and batch size of 64. The models are also fitted over a variable number of epochs ranging from 10 to 2000. Then a suitable value of epoch is chosen for the model corresponding to each monitoring site. For forecasting DO at a particular time of the day, the DO values measured at the same time for the past n days is taken as input. The performance of the models is also checked by considering a lag of 1 to 5 days as input for each monitoring station. After finding a suitable lag value for each monitoring station, a corresponding window of n previous DO values is given as input for the final building of the model.

The demonstrated model makes use of the LSTM layer as the first hidden layer. The window containing the sequential DO values are first given to the LSTM layer as input. This is followed by the attention layer, which assigns weights to the most relevant hidden states of the previous LSTM layer. The most relevant features generated is then passed to the fully connected dense layer. In the end, an output layer is added, which returns a single forecasted value.

The experiments demonstrate the LSTM combined with attention mechanism performed drastically better than the simple RNN, LSTM and bidirectional LSTM. Moreover, to compare and demonstrate the effectiveness of models, we employed root mean squared error (RMSE), coefficient of determination (R2), and mean absolute error (MAE). Table 2 shows the RMSE values while considering different number of DO values in the input window. The lower values of RMSE and MAE and higher R2 values indicate better models. Table 3 summarizes the results corresponding to RMSE, MAE and R2 for all the models thereby showing that the results obtained by combining LSTM with attention mechanism are significantly better as compared to the other models. The values also indicate that LSTM performs almost similarly or marginally better than the RNN model. Whereas the bidirectional LSTM gives significantly better results as compared to basic RNN and LSTM models. Further, attention combined with LSTM further outperforms all the models.

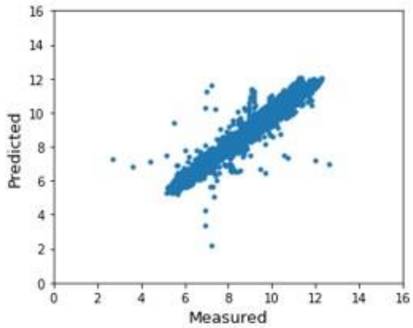
Table 2. RMSE of LSTM combined with attention model making use of different number of days as lags for River Ganga monitoring stations.

Station/ Number of days considered as lags	1 Day	2 Days	3 Days	4 Days	5 Days
Ghatiya Ghat bridge, Farrukhabad	0.344	0.343	0.345	0.348	0.347
Manimau bridge, Kannauj	1.771	1.762	1.752	1.758	1.757
Bithoor	0.817	0.823	0.849	0.850	0.853
Fatehpur (Asni Village)	1.276	1.262	1.310	1.257	1.320
Bridge in Varanasi	1.124	1.113	1.088	1.102	1.088

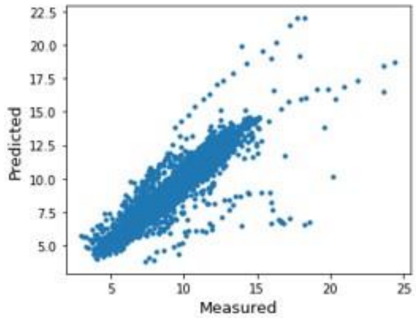
Figure 3 shows the scatter plots for all the stations for the testing phase of the dataset. The x-axis and the y-axis correspond to the measured and predicted values of DO respectively. Figure 4 shows the measured and predicted values of the DO during the testing phase of the model for the month of April 2021.

Table 3. Performance evaluation of models for River Ganga monitoring stations for Train and Test sets

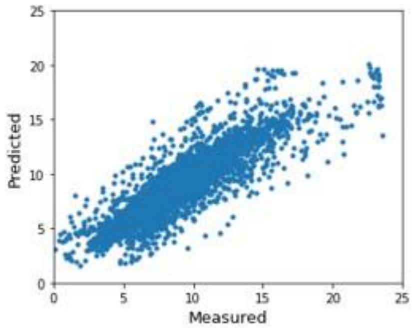
Stations	Models	RMSE		MAE		R^2	
		Train	Test	Train	Test	Train	Test
Ghatiya Ghat Bridge, Farrukhabad	RNN	0.630	0.464	0.337	0.293	0.832	0.904
	LSTM	0.630	0.464	0.338	0.295	0.832	0.904
	Bidirectional LSTM	0.458	0.344	0.226	0.198	0.911	0.947
	LSTM with Attention	0.457	0.343	0.224	0.197	0.911	0.947
Manimau Bridge, Kannauj	RNN	1.854	2.381	1.170	1.570	0.481	0.476
	LSTM	1.852	2.385	1.176	1.573	0.482	0.474
	Bidirectional LSTM	1.305	1.797	0.694	1.066	0.743	0.701
	LSTM with Attention	1.299	1.752	0.680	1.044	0.746	0.716
Bithoor	RNN	1.158	1.191	0.648	0.636	0.785	0.714
	LSTM	1.130	1.179	0.630	0.620	0.796	0.720
	Bidirectional LSTM	0.847	0.821	0.439	0.416	0.885	0.864
	LSTM with Attention	0.843	0.817	0.435	0.414	0.886	0.865
Fatehpur (Asni Village)	RNN	2.018	1.830	1.304	1.172	0.540	0.610
	LSTM	1.994	1.833	1.274	1.148	0.550	0.609
	Bidirectional LSTM	1.315	1.274	0.765	0.755	0.804	0.811
	LSTM with Attention	1.286	1.257	0.732	0.723	0.813	0.816
Bridge in Vara-nasi	RNN	2.170	1.760	1.452	1.172	0.587	0.794
	LSTM	2.159	1.732	1.438	1.144	0.591	0.801
	Bidirectional LSTM	1.494	1.089	0.920	0.638	0.804	0.921
	LSTM with Attention	1.480	1.088	0.912	0.639	0.808	0.921



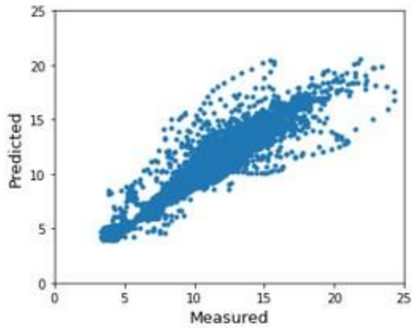
(a) Ghatiya Ghat bridge, Farrukhabad



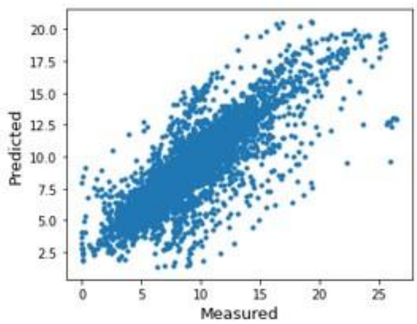
(b) Manimau bridge, Kannauj



(c) Bithoor

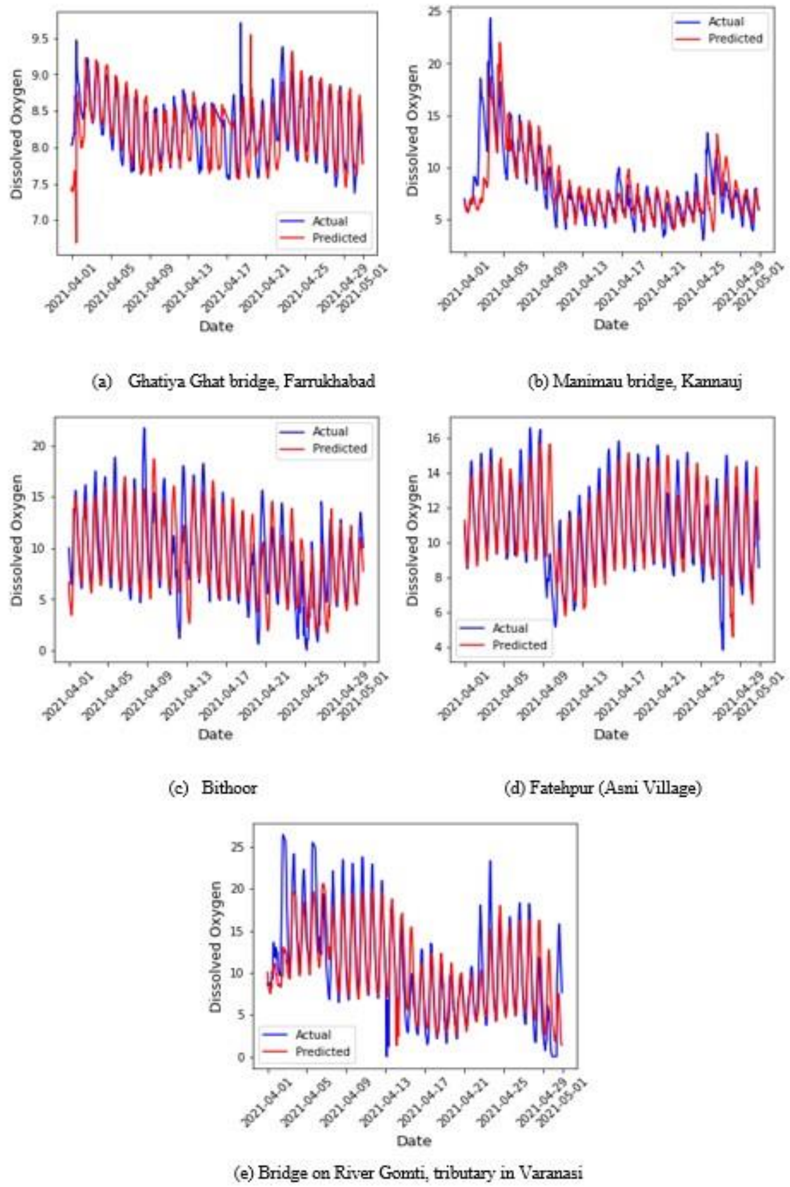


(d) Fatehpur (Asni Village)



(e) Bridge on River Gomti, tributary in Varanasi

Fig. 3. Scatter plots for the test dataset



(a) Ghatiya Ghat bridge, Farrukhabad

(b) Manimau bridge, Kannauj

(c) Bithoor

(d) Fatehpur (Asni Village)

(e) Bridge on River Gomti, tributary in Varanasi

Fig. 4. Plots showing measured and predicted Dissolved Oxygen for the test dataset

5 Conclusion

Managing water resources is one of the significant challenges for governments worldwide. Accurate long-term forecasting of water quality parameters like Dissolved Oxygen can help control the degree of pollution by encouraging the officials to make suitable policies and take precautionary actions.

Deep learning models like LSTM have been used successfully in several fields to forecast electricity demand [17][18], air pollution [19], rainfall [20], etc. Combining LSTM with the attention mechanism can significantly enhance its performance by assigning more weights to certain relevant hidden states for DO forecasting. The results suggest that the suggested temporal attention-based LSTM models perform better than the traditional standalone RNN, LSTM and bidirectional LSTM models. Further, the demonstrated model takes only previous DO values as input features and would therefore be helpful in situations when multivariate datasets are not available. Additionally, because the exhibited model simply relies on historical data, it is simple to extend it to any monitoring station.

References

1. Singh, K.P., Basant, A., Malik, A., Jain, G., 2009., "Artificial neural network modeling of the river water quality— a case study", Ecological Modelling, 220, 888–895. doi:10.1016/j.ecolmodel.2009.01. 004.
2. Ruben G. B. et al., "Application and Sensitivity Analysis of Artificial Neural Network for Prediction of Chemical Oxygen Demand", Water Resource Management (2018) 32:273–283 DOI 10.1007/s11269-017-1809-0
3. I. S. Yeon, J. H. Kim & K. W. Jun (2008), "Application of artificial intelligence models in water quality forecasting", Environmental Technology, 29:6, 625-631, DOI: 10.1080/09593330801984456
4. V. Rankovic, J. Radulovic, I. Radojevic, A. Ostojic, L. Comic, "Neural network modeling of dissolved oxygen in the Gruza reservoir, Serbia", Ecol. Model., 221 (8) (2010), pp. 1239-1244
5. S. Emamgholizadeh, H. Kashi, I. Marofpoor, E. Zalaghi, "Prediction of water quality parameters of Karoon River (Iran) by artificial intelligence-based models", Int. J. Environ. Sci. Technol. (2014) 11:645–656
6. Sarkar A., Pandey P. "River water quality modelling using artificial neural network technique", Aquatic Procedia, 4 (2015), pp. 1070-1077
7. M.J. Alizadeh, M.R. Kavianpour, "Development of wavelet-ANN models to predict water quality parameters in Hilo Bay, Pacific Ocean", Mar. Pollut. Bull., 98 (1) (2015), pp. 171-178
8. Csábrági, A., Molnár, S., Tanos, P., Kovács, J., 2017. "Application of artificial neural networks to the forecasting of dissolved oxygen content in the Hungarian section of the river Danube". Ecol. Eng. 100, 63–72. https://doi.org/10.1016/j.ecoleng.2016.12. 027.
9. Zou, Q., Xiong, Q., Li, Q. et al. A water quality prediction method based on the multi-time scale bidirectional long short-term memory network, Environmental Science and Pollution Research 27, 16853–16864 (2020). https://doi.org/10.1007/s11356-020-08087-7

10. Z. Li, F. Peng, B. Niu, G. Li, J. Wu, Z. Miao, "Water quality prediction model combining sparse auto-encoder and LSTM network", IFAC-PapersOnLine, 51 (17) (2018), pp. 831-836
11. Bahdanau, D., Cho, K., Bengio, Y., 2015. "Neural machine translation by jointly learning to align and translate". In: ICLR.
12. Z. Niu, Z. Yu, W. Tang, Q. Wu, M. Reformat, "Wind power forecasting using attention-based gated recurrent unit network Energy", 196 (2020), p. 117081
13. Ding Y.K., Zhu Y.L., Feng J., Zhang P.C., Cheng Z.R., "Interpretable spatio-temporal attention LSTM model for flood forecasting", Neurocomputing, 403 (2020), pp. 348-359
14. Haixiang Zang, Ruiqi Xu, Lilin Cheng, Tao Ding, Ling Liu, Zhinong Wei, Guoqiang Sun, " Residential load forecasting based on LSTM fusing self-attention mechanism with pooling", Energy, Volume 229, 2021, 120682, ISSN 0360-5442, https://doi.org/10.1016/j.energy.2021.120682.
15. Y. Liu, Q. Zhang, L. Song, Y. Chen, "Attention-based recurrent neural networks for accurate short-term and long-term dissolved oxygen prediction", Comput. Electron. Agric., 165 (2019), Article 104964, 10.1016/j.compag.2019.104964
16. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Computation 9(8):1735–1780
17. F.F. Guo, L.P. Li, C.H. Wei, "Short term load forecasting based on phase space reconstruction algorithm and bi-square kernel regression model", Appl. Energy, 224 (2018), pp. 13-33
18. T.Y. Kim, S.B. Cho, "Predicting residential energy consumption using CNN-LSTM neural networks", Energy, 182 (2019), pp. 72-81
19. Chang Y.S., Chiao H.T., Abimannan S., Huang Y.P., Tsai Y.T., Lin K.M., "An LSTM-based aggregated model for air pollution forecasting", Atmos. Pollut. Res., 11 (2020), pp. 1451-1463
20. Y. Xu, C. Hu, Q. Wu, S. Jian, Z. Li, Y. Chen, G. Zhang, Z. Zhang, S. Wang, "Research on Particle Swarm Optimization in LSTM Neural Networks for Rainfall-Runoff Simulation", J. Hydrol. (2022), Article 127553, 10.1016/j.jhydrol.2022.127553