A Multi-factor Water Quality Prediction Method Based on Wavelet Transform and LSTM

Mingxia Yang[1], Lianghuai Tong[2], Aiping Xia[2], and Kai Fang[1](✉)

[1] College of Electrical and Information Engineering, Quzhou University, Quzhou, Zhejiang, China
Kaifang@ieee.org

[2] Quzhou Academy of Metrology and Quality Inspection, Quzhou, Zhejiang, China

Abstract. Water resources are an important natural resource for mankind. Protecting water resources is the key to maintaining human survival and development. Water quality is affected by many factors, including physical and chemical factors, so the accuracy of traditional water quality prediction methods is not yet satisfactory. In order to improve the accuracy of water quality prediction, this paper proposes a multi-factor water quality prediction method based on wavelet transform and LSTM (WT-LSTM). Firstly, we select multi-featured factors in the water quality data as inputs, then, noise reduction is performed on each original feature based on wavelet decomposition, and finally, the noise reduced data are input into LSTM for estimation. The experimental results show that the prediction performance of WT-LSTM is better than the original LSTM prediction model, and the multifactor prediction is better than the single-factor method. The final experimental coefficient of determination is 0.9650, which is higher than the comparison model.

Keyword: Water quality prediction · Wavelet transform · LSTM · Multi-factor

1 Introduction

Water represents an indispensable resource for humans and other organisms to survive on the earth [1, 2]. Due to economic development, many industrial wastewaters are produced, which seriously damages the ecological environment and ecological functions of water. With population growth and economic development, China's water consumption is increasing, but serious water pollution has reduced the utilization of various freshwater resources.

Water quality prediction is built on the historical data from previous monitoring, through the water quality prediction model to predict the future trend of water quality data [3]. In the prevention and control measures of water resources, water quality prediction has become more and more important [4–6]. Effective and accurate water quality prediction data can monitor the trend of water quality changes in real time and take countermeasures in advance, effectively avoid and prevent water quality deterioration, improve water pollution control, and promote ecological restoration, the

V. C. M. Leung et al. (Eds.): Qshine 2023, LNICST 574, pp. 130–144, 2024.
https://doi.org/10.1007/978-3-031-65123-6_10

future development of the water environment is of great significance [7–9]. The most common indicators of water quality prediction are PH value, dissolved oxygen, total phosphorus, ammonia nitrogen, biological oxygen demand, chemical oxygen demand, and some physical indicators, such as water temperature, conductivity, and turbidity. At the beginning of the water quality prediction research, it is the dissolved oxygen (DO) as the center of research, to study the relationship between the consumption rate of DO and BOD in the water body, and the relationship between the oxygen deficit value in the water body and the rate of oxygen entering the water body through the water surface [3, 10].

Machine learning is employed for water quality prediction research, and better results are achieved [11–15]. Li et al. [16] proposed a water quality prediction model based on multi-source data machine learning for the shortage of meteorological, hydrological, and socio-economic factors considered in traditional water quality prediction, and experimentally proved that machine learning can integrate meteorological, socio-economic and other multi-source and different spatio-temporal data to complete hierarchical water quality prediction, and the best performance of the prediction model based on random forest. Lu et al. [11] proposed two new hybrid decision tree based machine, using these two prediction models for water quality prediction in Tarakin River basin, and the results show that the hybrid decision tree based machine learning model predicts better than the traditional model. Juna et al. [17] proposed a nine-layer multilayer perceptron with K-nearest neighbor algorithm together with the K-nearest neighbor algorithm for dealing with the missing value problem, and the experimental results demonstrate that the proposed 9-layer MLP model has excellent water quality prediction accuracy up to 0.99 under the KNN algorithm.

With the development of deep learning, Neural networks are increasingly used in a wide range of applications [18–22], and many literature experiments have demonstrated the effectiveness of neural networks in water quality prediction [22–24]. Tu et al. [25] proposed a water quality prediction model based on deep learning gated recurrent neural network in order to optimize and improve the accuracy of water quality prediction model, and proved that the water quality prediction model based on deep learning significantly improved the prediction accuracy than the traditional water quality prediction model. Wang et al. [26] used a neural network technology model for water quality prediction, and concluded that the BP neural network total nitrogen prediction model with gray correlation degree can effectively predict water quality changes, but it needs to be upgraded and optimized under the conditions of multiple influencing factors. Ali et al. [27] proposed an enhanced wavelet noise reduction technique based on neuro-fuzzy inference system. The enhanced wavelet noise reduction technique of water quality prediction model based on neuro-fuzzy inference system is verified to have significant improvement over the previous water quality models.

In order to improve the LSTM prediction accuracy, the original data can be processed by noise reduction first [28, 29]. Chen et al. [30] considering the problem of large-scale continuous missing data, a TrAdaBoost-LSTM model is proposed and the effectiveness of the method is experimentally verified. Sun et al. [31] proposed a wavelet decomposition based LSTM model based on LSTM model for four water quality indicators, which has higher prediction accuracy and better effect than the traditional LSTM model.

In the prediction process, it is divided into single-factor and multi-factor based prediction according to the dimensionality of eigenvalues. Shi et al. [32] established a water quality prediction model based on wavelet analysis with long and short memory neural network, which decomposes water quality data into high-frequency and low-frequency signals, and then inputs them into the LSTM model for single-factor water quality prediction. It is noted that wavelet analysis method can better capture the characteristics of water quality data, and its combination with LSTM water quality prediction model can significantly improve the accuracy of water quality prediction. However, the study used a single factor prediction that does not reflect overall water quality conditions.

In terms of multi-factor prediction, Wang [33] proposed an improved gray correlation analysis algorithm IGRA based on the traditional gray correlation analysis algorithm, and the correlated water quality indicators selected based on IGRA were input into LSTM together with the target water quality indicators for water quality prediction, and the experimental results showed that the water quality prediction method based on IGRA-LSTM can make full use of the multiple correlations of water quality indicators, and the accuracy of the prediction model using the correlated indicators as the input of multiple features was improved. The accuracy of the prediction model using correlated indicators as multi-feature input has been improved. Liu [34] proposed a multi-factor water quality prediction model based on LSTM, using K-Similarity noise reduction method to reduce the noise of the model input data and improve the model prediction performance.

In order to improve the accuracy of water quality prediction, this paper uses the multi-factor characteristics of water quality data, combining wavelet decomposition and LSTM to study water quality prediction methods. The main contributions of this paper are as follows.

(1) Summarize the current status of water quality prediction, summarize and analyze the research around deep learning models in water quality prediction, and compare the advantages and disadvantages of single-factor and multi-factor water quality prediction.
(2) Establish a wavelet transform-based LSTM water quality prediction model, compare the accuracy difference between the wavelet transform-based LSTM water quality prediction model and the traditional LSTM model in predicting water quality data, train and test the two sets of models, and finally compare and analyze the prediction results.
(3) Experimentally compare and analyze the algorithm on datasets with different time spans, and use wavelet transform for noise reduction on the basis of multi-factor features to further improve the effectiveness of the algorithm.

2 Data Sources and Prediction Framework

2.1 Data Sources

Water environment is a complex gray system, water quality indicators under the influence of various factors, non-linear changes over time, with complex multivariate correlation. Water quality contains many elemental content, such as PH, total phosphorus, dissolved oxygen, total nitrogen, nitrate nitrogen, potassium ions, sodium ions, etc. These elements will also change with the physical environment changes, physical and chemical factors combined with each other to affect the prediction of water quality data. When the dissolved oxygen in the water is below a certain threshold, the fish and shrimp in the water will die of lack of oxygen.

The data set of this paper is the monthly observation data of 8 monitoring stations in Taihu Lake from 2000 to 2006, total 672 items, provided by the National Earth System Science Data Sharing Service Platform. The physical water quality indicators include transparency, rainfall, suspended solids, and chemical water quality indicators include PH, total nitrogen, total phosphorus, potassium permanganate index, dissolved oxygen, ammonia nitrogen, and chloride. Among the chemical indicators, dissolved oxygen and total nitrogen are two principal water quality pollution indicators.

The water quality data of Taihu Lake from January 2001 to December 2006 were used, and the data from testing station TH01 is taken here, with a total of 190 records. All the above data are relatively complete, so no vacant values were filled, and only normalization was performed. In the water quality data, the total nitrogen content is not consistent with the normal distribution, so the maximum-minimum normalization is used. Because of the small amount of data, the window size was chosen as 3. The data of the first 3 months were used to predict the data of the next month.

2.2 Forecasting Framework

2.2.1 Wavelet Transform

The LSTM model has strong advantages for predicting time series data, but it is more difficult to obtain data patterns for data with high variability and complexity, and the prediction results obtained will be poor. Wavelet analysis is a mathematical theory and method that has emerged in the last decade or so. Wavelets can perform effective time-frequency decomposition of signals, which is a flexible time-frequency analysis technique that can separate real signals from excess noise. Figure 1 shows the general form of wavelet function.

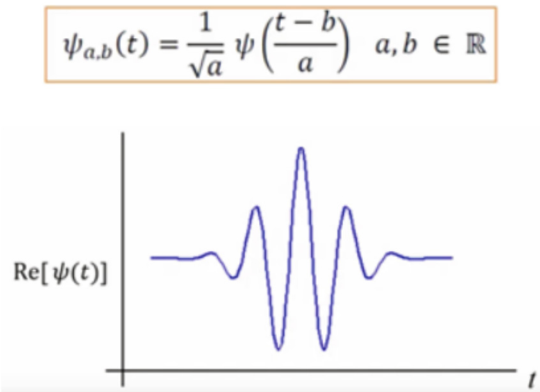

Fig. 1. General form of wavelet function

For the amount of mutation in the signal, there is a Gibbs effect in the Fourier transform, and the mutation signal is not well fitted with the infinitely long trigonometric function, while the attenuated wavelet can be better fitted to the mutation signal. As shown in Fig. 2, for the mutation signal, the coefficients are not zero only when the wavelet function and the signal mutation overlap.

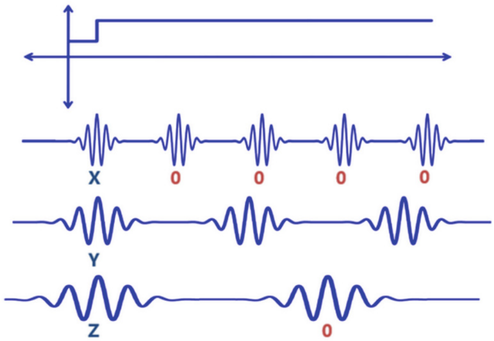

Fig. 2. Wavelet function fitting mutation signal

After wavelet decomposition, the result of each layer of decomposition is that the low-frequency signal obtained from the previous decomposition is decomposed into two parts: low-frequency and high-frequency. After N layers of decomposition, the source signal X is decomposed as follows: X = D1 + D2 +… + DN + AN + DN + AN where D1, D2,… DN are the high-frequency signals obtained from the first, second and N-layer decompositions, and AN is the low-frequency signal obtained from the N-layer decomposition.

The reconstructed data series obtained after wavelet transform possesses the ability to characterize the local features of the signal in both time and frequency domains, and it is easier to detect singularities or transients of the signal than the original data series signal, which has obvious advantages for analyzing and processing non-smooth time series. In this paper, db10 is selected as the basis wavelet for the reason that db10 is more suitable for the decomposition of relatively smooth data sets.

2.2.2 LSTM-Based Water Quality Prediction Model

This paper uses LSTM model for water quality prediction, which is faster and easier to converge to the optimal solution than traditional neural networks when dealing with time series problems, and is very suitable for dealing with time series like water quality indicators.

Let the water quality indicators of the n chronological arrangement of the historical data values constitute a time series of $x_i(n) = [x_i(1), x_i(2)...x_i(m)..x_i(n)]$, i represents the name of different water quality indicators, such as x_i represents dissolved oxygen, then $x_i(n)$ represents the n chronological historical monitoring values of dissolved oxygen, $x_i(m)$ represents the observed values of dissolved oxygen at the mth moment. Set m moment for the current moment, water quality indicators x_i in the previous d moments to m moments of the data series, where d is used to indicate the size of the sliding window. When using the features for prediction, based on the size of d, the data of the water quality indicator x_i will be passed to the input layer of the LSTM water quality prediction model, so the predicted value of the next moment of the sequence $\widehat{x}_i(m+1)$.

Each implicit layer neuron contains an input gate I, an output gate O, a forgetting gate F, and the current state $c(m)$. The gating mechanism implements the selective memory function of the LSTM, which makes the LSTM more suitable than the traditional neural network to handle the temporal prediction problem. The computation process of each part of the implicit layer neuron is as follows.

$$\mathrm{h_j(m) = H(m)(W_{ih}X_i(m) + W_{hh}h_{j\text{-}1}(m) + b_h)} \tag{1}$$

$$\widehat{x}_i(m+1) = \sigma(W_{ho}H(m) + b_y) \tag{2}$$

The calculation process of the forgetting gate F:

$$F = \sigma(W_{fi}x_i(m) + W_{fc}c_m + W_{fh}h(m-1) + b_f) \tag{3}$$

Calculation procedure for input gate I:

$$I = \sigma(W_{Ii}x_i(m) + W_{Ic}c_{m-1} + W_{\mathrm{I}h}(m-1) + b_I) \tag{4}$$

Current status $c(m)$ Update the calculation process:

$$c_m = F * c_{m-1} + I * g(W_{ci}x_i(m) + W_{ch}h(m-1) + W_{cc}c_{m-1} + b_c) \tag{5}$$

The procedure for calculating the output gate O:

$$\mathrm{O} = \sigma(W_{oi}x_i(m) + W_{hh}h(m-1) + W_{oc}c_{m-1} + b_o) \tag{6}$$

The output of the implicit layer at moment m is

$$h(m) = O * P(c_{m-1}) \tag{7}$$

Based on the LSTM model to predict water quality indicators, the size of the sliding window d is set and the sequence is used to predict the value of the next moment. When the sliding window d is 3, it means that using the data of the first three moments to predict the data that may be generated in the next moment, for example, using the data of January, February and March will be able to predict the data model of the 4th month, using the data of April, May and June will be able to predict the data model of the 7th month, and so on.

2.2.3 Process Design

Based on the LSTM model, wavelet transform is added to the original data for signal processing to improve the accuracy of the prediction model. Based on wavelet transform and LSTM model of water quality prediction flow chart is shown in Fig. 3.

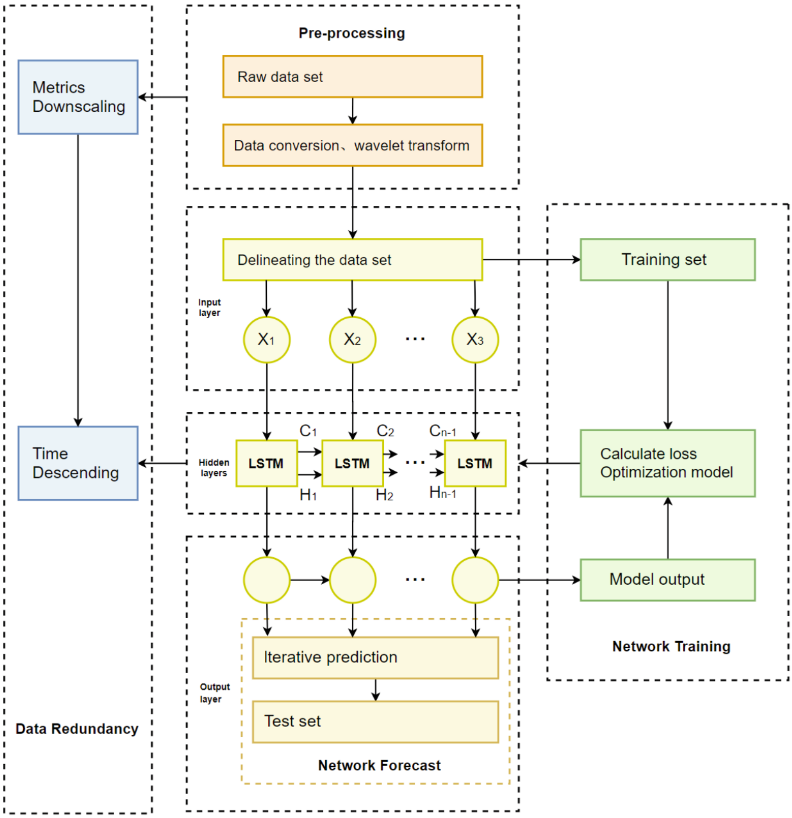

Fig. 3. System flow and framework

The steps to achieve the prediction are as follows.

Step 1. The collected water quality index data were normalized by using mean smoothing method to reduce noise.
Step 2. The sample data are uniformly divided into the first 80% as training data and the second 20% as test data.
Step 3. Using the training data as the sample input for training the WT-LSTM neural network model, the following two steps are performed on the model.

(1) Select “db10” as the base wavelet, and perform 3rd order wavelet decomposition on the data to obtain the low frequency signal A3 and high frequency signals D1, D2 and D3;
(2) Using LSTM to predict A3, D1, D2, and D3 respectively.

Step 4. Use the test data as the input sample of WT-LSTM model, output the prediction accuracy of the model and compare the error with the comparison test model.

3 Experimental Results and Analysis

3.1 Experimental Environment and Parameter Settings

PH, total nitrogen, total phosphorus, ammonia nitrogen, nitrite nitrogen, nitrate nitrogen, potassium ion, sodium ion, calcium ion multi-factor data were selected as the eigenvalue input for wavelet transform based LSTM water quality prediction, and the data were pre-processed separately using the wavelet transform method first. Among them, 190 data values of total nitrogen were selected as the eigenvalue input.

In the LSTM model prediction, the number of nodes in the input layer is determined by the size of the sliding window d, and the number of nodes in the output layer is 1. Because the number of nodes in the hidden layer is determined by relying on continuous experimental results, in order to make the LSTM-based numerical prediction model more accurate, through training, comparing the results of different parameters, the number of hidden layer neurons is finally set to 12, the activation function is After several tests, the number of iterations epochs was 150, the number of single training samples batch size was set to 16, and the learning rate was set to 0.005.

In this study, the mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination (R^2) are used to measure the prediction effect of the model. MAE and RMSE are measures of deviation between the true and predicted values, which are related to the magnitude, and the smaller the value indicates the better prediction effect of the model, and the closer R^2 is to 1 indicates the better fitting ability of the model, which can be compared with the goodness of fit of models of different magnitudes.

3.2 Analysis of Experimental Results

3.2.1 LSTM-Based Prediction

Total nitrogen is better predicted in a sliding window of 3, the number of nodes in the implied layer is 30, the epoch is 150, MAE, RMSE, MAPE are less than 0.1, and R^2 is 0.9446. The loss function starts to fluctuate more frequently after 20 iterations, which can be adjusted by modifying the learning rate (Fig. 4).

It is found that the number of implicit layer nodes and iterations are not better under the same window, and different relative parameter values are obtained under different windows. In Fig. 5, it can be seen that the error at the extreme value, especially at the minimum value, is larger, and some adjustments can be made to the parameters.

3.2.2 Prediction Results Based on WT-LSTM

Wavelet transform was done for one of the dependent variables, TN, and the initial image of the TN content is shown in Fig. 6.

The waveform of TN is decomposed into d1, d2, d3, a3 respectively. A three-layer wavelet decomposition is performed and the decomposed waveform is shown in Fig. 7.

Finally, the TN image and data after wavelet processing are output. Compared with the original image, the wavelet transformed image is clearer and smoother.

Through wavelet decomposition, reconstructed data sequences can be obtained which are able to characterize the local features of the signal in both time and frequency domains, thus making it easier to detect real-time states or singularities of the

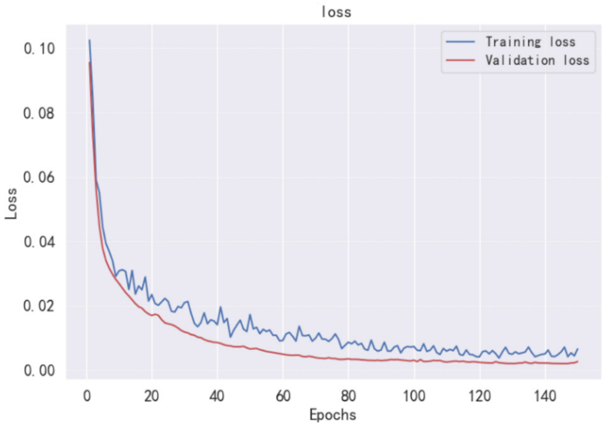

Fig. 4. Loss function

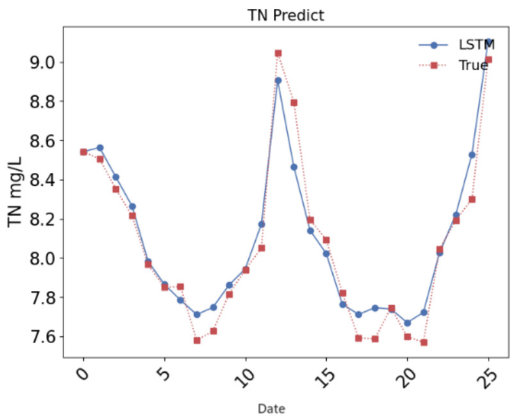

Fig. 5. Comparison of true and predicted values

signal. When analyzing and processing non-stationary time series data, LSTM does not predict mutations well. Wavelet transform has the ability to characterize local features, so it can detect real-time states and singularities, which has obvious advantages (Fig. 8).

The wavelet transform was applied to all dependent and independent variables and then put into the LSTM neural network model for training, but the resulting images are as shown in Fig. 9. The difference between the wavelet predicted data and the true value is very large. We found that because the original data was wavelet transformed and then put into the LSTM neural network model for training. The original data is the original data for LSTM neural network training to produce the predicted value and the real value comparison, so the wavelet transformed data then compared with the original data will have a large error and should be compared with the original data after wavelet transform.

The original data with wavelet transform is also put into the image, as shown in Fig. 10. It can be seen that the accuracy of the data that have undergone wavelet transform

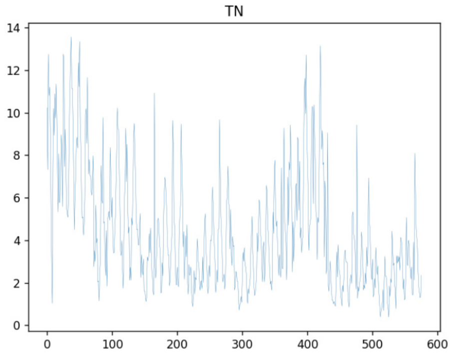

Fig. 6. The original TN data

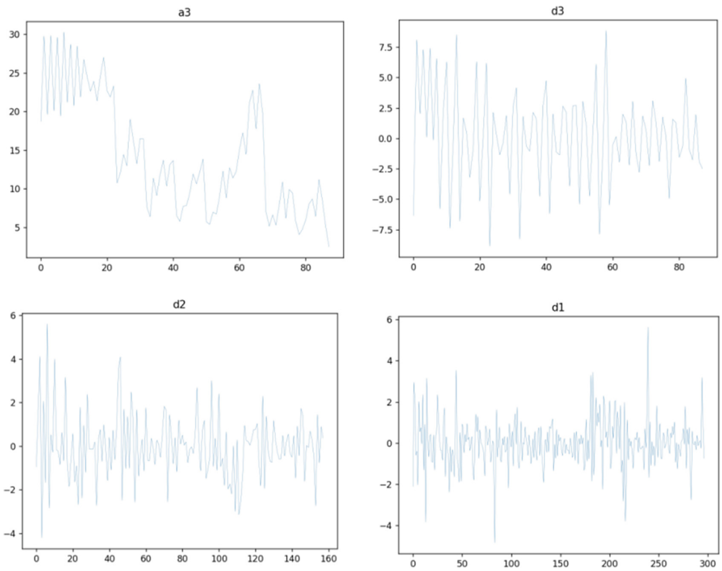

Fig. 7. Three-layer wavelet decomposition of TN

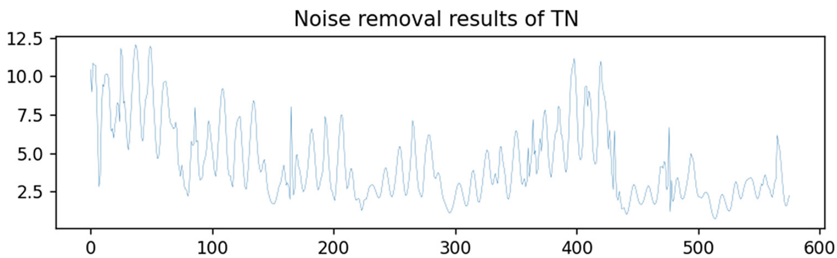

Fig. 8. TN after wavelet transform

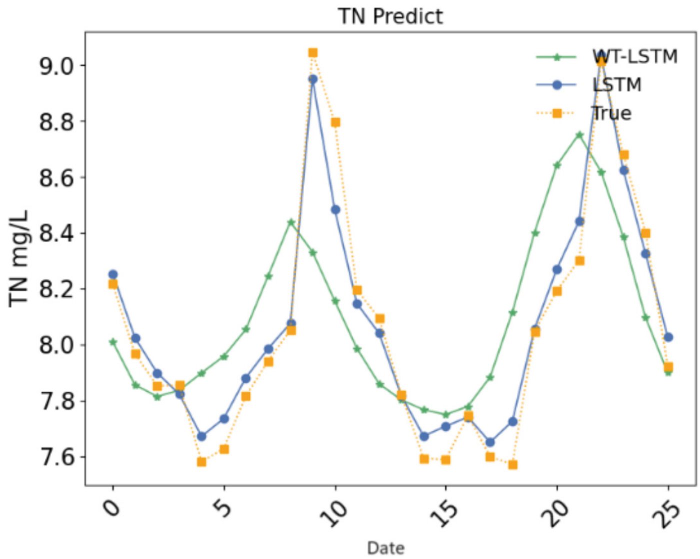

Fig. 9. Prediction results

is higher at the minimum value and relatively higher at the maximum value compared to the predicted values of LSTM without wavelet transform.

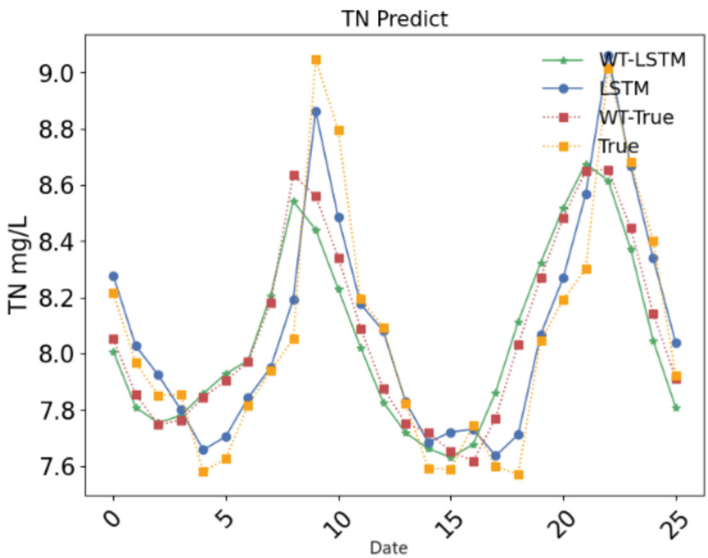

Fig. 10. Wavelet transform comparison experiment

In order to use the same reference data to measure the model effect, the initial data was wavelet transformed as the true value, and then the data was put into the LSTM model for testing to get the predicted value (LSTM), and then the predicted value 1 was wavelet transformed and compared with the true value. The results are shown in Fig. 11. It can be seen that the data without wavelet transform has a large error at both the maximum and minimum values, while the wavelet transformed data can better match the true value at both the maximum and minimum values. After wavelet decomposition and then using the long-short memory neural network model for time series prediction more accurately reflects the overall trend and grasps the details of changes.

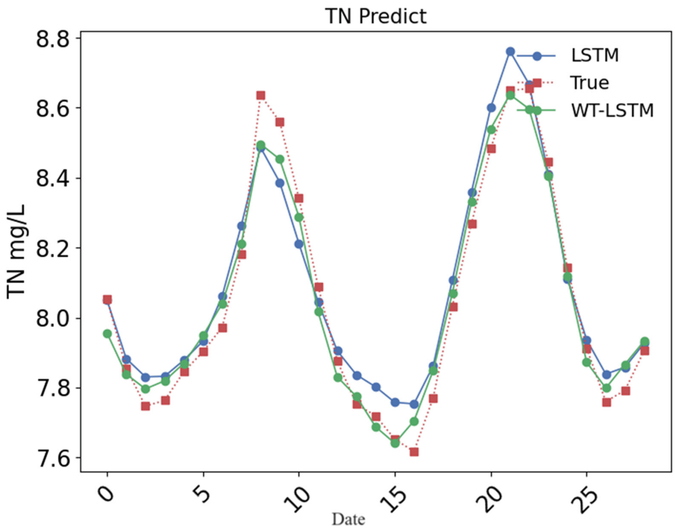

Fig. 11. Comparison of prediction results

As shown in Table 1, after scoring the models, it can be seen that the prediction model incorporating the wavelet transform is more accurate than the regular LSTM data in all metrics.

The experimental results prove that LSTM neural network has a strong advantage in predicting time series data, but for the data with high complexity and high frequency of change, it is difficult for a single LSTM prediction method to obtain the change rule of the data, which leads to poor simulation and prediction results. Wavelet decomposition can decompose the information of different frequency bands in the original data, greatly reducing the complexity of the data, and then predict these data separately to improve the prediction accuracy.

Table 1. Evaluation and comparison of Models

Models	MAE	MAPE	RSME	R^2
LSTM	0.0715	0.0990	0.0088	0.9446
WT-LSTM	0.0519	0.0064	0.0597	0.9650

4 Conclusion

In this paper, we studied the water quality prediction based on time series analysis, and proposed a multifactor water quality prediction method integrating wavelet transform and LSTM for the implied complexity and high-frequency variations of water quality data, and used different waveform decomposition to improve the prediction accuracy. The experimental results show that the results without wavelet transform data have larger errors at the maximum and minimum values, while the results with wavelet transform data can better match the real values at both the maximum and minimum values. Multi-factor can fully utilize the multiple correlations of water quality indicators, which can better reflect the overall water quality condition.

In terms of the LSTM-based water quality prediction method, this paper has achieved certain research results, but the method still has some areas that need to be improved and perfected. Due to the complex structure of LSTM neurons, the training time of the prediction model is long, and it can be considered to improve its neuron structure to shorten the training time, thus shortening the training period. How to improve the LSTM to better reflect the intrinsic mechanism of water quality change to further improve the prediction accuracy still needs to be explored.

References

1. Rajaee, T., Khani, S., Ravansalar, M.: Artificial intelligence-based single and hybrid models for prediction of water quality in rivers: a review. Chemom. Intell. Lab. Syst. **200**, 103978 (2020)
2. Li, D., Sun, Y., Sun, J., Wang, X., Zhang, X.: An advanced approach for the precise prediction of water quality using a discrete hidden Markov model. J. Hydrol. **609**, 127659 (2022)
3. Zhang, Y., et al.: Accurate prediction of water quality in urban drainage network with integrated EMD-LSTM model. J. Clean. Prod. **354**, 131724 (2022)
4. HaRa, J., Mamun, M., An, K.G.: Ecological river health assessments using chemical parameter model and the index of biological integrity model. Water **11**(8), 1729 (2019)
5. Woznicki, S.A., Nejadhashemi, A.P., Ross, D.M., et al.: Ecohydrological model parameter selection for stream health evaluation. Sci. Total. Environ. **511**, 341–353 (2015)
6. Chen, Y.M., Xia, J.H., Cai, W.W., et al.: Three-phase-based approach to develop a river health prediction and early warning system to guide river management. Appl. Sci. **9**(19), 4163 (2019)
7. Jiang, Y., Li, C., Sun, L., Guo, D., Zhang, Y., Wang, W.: A deep learning algorithm for multi-source data fusion to predict water quality of urban sewer networks. J. Clean. Prod. **318**, 128533 (2021)
8. Zhang, Y., Li, C., Jiang, Y., Zhao, R., Yan, K., Wang, W.: A hybrid model combining mode decomposition and deep learning algorithms for detecting TP in urban sewer networks. Appl. Energy **333**, 120600 (2023)
9. Wang, X., Li, Y., Qiao, Q., Tavares, A., Liang, Y.: Water quality prediction based on machine learning and comprehensive weighting methods. Entropy **25**, 1186 (2023)
10. Jia, X., et al.: Physics Guided Recurrent Neural Networks for Modeling Dynamical Systems: Application to Monitoring Water Temperature and Quality In Lakes. ArXiv, abs/1810.02880 (2018)
11. Lu, H., Ma, X.: Hybrid decision tree-based machine learning models for short-term water quality prediction. Chemosphere **249**, 126169 (2020)

12. Wang, Y., Zhou, J., Chen, K., Wang, Y., Liu, L.: Water quality prediction method based on LSTM neural network. In: 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE) (2017)
13. Malek, N.H., Wan Yaacob, W.F., Md Nasir, S.A., Shaadan, N.: Prediction of water quality classification of the Kelantan river basin, Malaysia, Using Machine Learning Techniques. Water **14**(7), 1067 (2022)
14. Hayder, G., Kurniawan, I., Mustafa, H.M.: Implementation of machine learning methods for monitoring and predicting water quality parameters. Biointerface Res. Appl. Chem. **11**(2), 9285–9295 (2020)
15. Kazakis, N.: Improving prediction of water quality indices using novel hybrid machine-learning algorithms. Sci. Total. Environ. **742**, 137612 (2020)
16. Zhou, J., Wang, J., Chen, Y., Li, X., Xie, Y.: Water quality prediction method based on multi-source transfer learning for water environmental IoT system. Sensors **21**, 7271 (2021)
17. Juna, A., Umer, M., Sadiq, S., Karamti, H., Eshmawi, A.A., Mohamed, A.E., Ashraf, I.: Water quality prediction using KNN imputer and multilayer perceptron. Water **14**, 2592 (2022)
18. Wang, T., Fang, K., Wei, W., Tian, J., Pan, Y., Li, J.: Microcontroller unit chip temperature fingerprint informed machine learning for IIoT intrusion detection. IEEE Trans. Industr. Inf. **19**, 2219–2227 (2023)
19. Mao, K., Xu, J., Jin, R., Wang, Y., Fang, K.: A fast calibration algorithm for non-dispersive infrared single channel carbon dioxide sensor based on deep learning. Comput. Commun. **179**, 175–182 (2021)
20. Hmoud Al-Adhaileh, M., Waselallah Alsaade, F.: Modelling and prediction of water quality by using artificial intelligence. Sustainability **13**, 4259 (2021)
21. Prasad, D.V.V., et al.: Analysis and prediction of water quality using deep learning and auto deep learning techniques. Sci. Total. Environ. **821**, 153311 (2022)
22. Liu, P., Wang, J., Sangaiah, A.K., Xie, Y., Yin, X.: Analysis and prediction of water quality using LSTM deep neural networks in IoT environment. Sustainability **11**, 2058 (2019)
23. Hu, Z., et al.: A water quality prediction method based on the deep LSTM network considering correlation in smart mariculture. Sensors **19**, 1420 (2019)
24. Hu, Y.K., Wang, N., Liu, S.H., Jiang, Q., Zhang, N.: Research on the application of time series model and LSTM model in water quality prediction. Small Microcomputer Systems **42**(08), 1569–1573 (2021)
25. Tu, J.C., Chen, C., Wang, J., Wang, Z., Ye, Q.: Research on water quality prediction model based on deep learning. Autom. Instrum. **34**(06), 96–100 (2019)
26. Wang, M., Zhang, L., Shan, W.: Research on water quality prediction model based on neural network. Autom. Expo **39**(12), 66–69 (2022)
27. Ali Najah Ahmed, A.N.A., et al.: Machine learning methods for better water quality prediction. J. Hydrol. **578**(C), 124084–124084 (2019)
28. Wang, T., Li, J., Wei, W., Wang, W., Fang, K.: Deep learning-based weak electromagnetic intrusion detection method for the zero touch industrial internet of things. IEEE Network **36**(6), 236–242 (2022)
29. Wei, S.K., Yang, H., Song, J.X., et al.: A wavelet-neural network hybrid modelling approach for estimating and predicting river monthly flows. Hydrol. Sci. J. **58**(2), 374–389 (2013)
30. Chen, Z., et al.: A transfer learning-based LSTM strategy for imputing large-scale consecutive missing data and its application in a water quality prediction system. J. Hydrol. **602**, 126573 (2021)
31. Sun, M., Wei, S., Wang, Y., Zhao, J., Yuan, M.: LSTM water quality prediction model based on wavelet decomposition. Comput. Syst. Appl. **29**(12), 55–63 (2020)
32. Shi, Q., Dong, Z., Luo, E., et al.: Research on the evaluation and prediction of water quality of inlet lake of Hongze Lake based on machine learning method. China Rural Water Conservancy Hydropower (12), 7 (2021)

33. Zhou, J., Wang, Y., Xiao, F., Wang, Y., Sun, L.: Water quality prediction method based on IGRA and LSTM. Water **10**(9), 1148–1159 (2018)
34. Liu, J., Zhuang, H., Tie, Z., et al.: K-Similarity noise reduction of LSTM neural network for water quality multi-factor prediction model. Comput. Syst. Appl. **2**, 7 (2019)