The Development of Humanities and Social Sciences, 2025, 1(1), 1000119
https://doi.org/10.71204/t8xden06

Factory Pollutant Discharge Data Flow Prediction Based On LSTM-Transformer

Hongyu Zhang [1, 2,*], Fujiang Yuan [1]

[1]School of Computer Science and Technology, Taiyuan Normal University, Jinzhong 030619, China.

[2]Department Electronic Engineering,TaiYuan institute of technology,Taiyuan 030008, China

* **Correspondence**:

Hongyu Zhang

zhy3090@163.com

Received: 6 March 2025/Accepted: 4 April 2025/Published online: 7 April 2025

Abstract

Aiming at the problems of complex feature extraction and insufficient ability to capture time series dependencies in the prediction of factory pollutant discharge data, this study proposes a hybrid deep learning model that integrates a Long Short-Term Memory (LSTM) network with a Transformer. Multi-dimensional time series data on factory pollutant discharge (such as hourly flow rate, pH value, ammonia nitrogen concentration, etc.) are collected via Internet of Things (IoT) devices. After standardization, local temporal features are extracted using the LSTM network, while global dependency relationships are captured through the Transformer's multi-head self-attention mechanism. Experiments are conducted on a public dataset, using root mean square error (RMSE), mean absolute error (MAE), and the coefficient of determination (R^2) as evaluation metrics to compare the traditional LSTM with the Transformer model. The results show that the LSTM−Transformer model achieves the best performance in predicting six types of pollutant discharge data (with RMSE reduced by 6.3%−12.1% and R^2 improved by 2.5%−5.2%), demonstrating its effectiveness in accurately capturing both long-term and short-term dependencies and providing robust support for real-time pollutant discharge early warning in smart factories.

Keywords: LSTM; Transformer; Factory Sewage Discharge

1. Introduction

The rapid economic development and continuous industrialization and urbanization in China have made the issue of water pollution increasingly prominent, severely restricting the country's high-quality economic and social development. On April 16, 2015, the State Council officially released the action guidelines for water pollution prevention and control called the Water

Pollution Prevention Action Plan , which points out that "comprehensive control of pollutant emissions and strict management of industrial pollution prevention" is necessary, indicating that water pollution mainly originates from industrial discharges. How to comprehensively control water pollution is an urgent problem that needs to be addressed.

In recent years, with the rapid development of technology, the application of Internet of Things(loT) (Stefano, P., Giacomo, P., & Alessandro, P. 2021) technology has become increasingly widespread. By placing loT devices at factory wastewater discharge points, vast amounts of data have been integrated into the internet via sensors. To mine valuable insights from the acquired data, researchers employ various data analysis techniques to process the numerous metrics produced by wastewater treatment facilities. With the advancements in artificial intelligence, some professionals have chosen to apply machine learning knowledge to wastewater treatment early warning systems. However, traditional machine learning algorithms have their limitations. They often rely heavily on the representation of data features (Joseph, G., Moyez, D., Paulo, N., & others. 2014), which are usually selected manually based on experience. This method complicates the extraction of important features and presents challenges since empirical selections do not always provide reliable guidance. This experience-based method of feature selection not only increases the difficulty of feature extraction but also limits the accuracy and generalizability of the models.

In this context, considering the application of deep learning to wastewater data analysis may be viable. Applying deep learning techniques to wastewater treatment is an emerging research direction, involving the collection, analysis, and processing of information such as temperature adjustments and emission standard settings. This requires an integration of knowledge from mathematics, computational science, and environmental engineering to interpret and analyze the implications of the data and its practical applications. By employing an interdisciplinary approach researchers can achieve a more accurate interpretation of data, optimize the wastewater treatment process, and ensure compliance with environmental requirements. The advantage of deep learning lies in its ability to automatically learn and identify complex patterns in data without relying on empirical feature selection. As a result, deep learning has found widespread applications in various fields such as image processing (David, D., & others. 2016)., speech recognition (LiMin, Z., & others. 2023), machine translation (Yuan, F., & Peng, Y. (2025)., and natural language processing (Baosong, Y., & others. 2021).. Additionally, deep learning has an important application scenario in smart factories, which encompass areas such as wastewater treatment warnings and intelligent equipment management. Smart factories aim to autonomously extract valuable information from vast amounts of data without human intervention, thereby enhancing the accuracy and efficiency of the system.

2. Related Work

In the field of wastewater treatment, numerous challenges remain. Many researchers have conducted considerable work related to this area. For instance, Liu Hui et al. proposed a sewage treatment method based on image processing technology.Through machine learning and artificial intelligence algorithms, the data are deeply analyzed to improve the intelligent level of sewage treatment (Hui, L., & Linguo, L. (2018). Maryam Mahjouri et al. proposed a decision support tool that combined fuzzy Delphi method with fuzzy Analytic Hierarchy Process (FAHP) to efficiently deal with complex decision problems, and verified the objectivity and practicability of the model integration. The study aims to identify key evaluation criteria and indicators for the selection of industrial wastewater treatment technologies from the perspective of sustainable development, based on expert opinions and questionnaires, and to conduct a practical evaluation using the Iranian steel industry as a case study. The research results provide reference for steel industry in selecting the best wastewater treatment technology (Mahjouri, M., & others. 2017). Taking coal chemical wastewater treatment as the research object, He Miao established a microwave-assisted numerical simulation algorithm for coal chemical wastewater treatment based on JAVA language, and compared the test results with the numerical simulation results to verify the effectiveness of the proposed numerical calculation method, and analyzed the influence of different pH values and H2O2 dosage on coal chemical wastewater treatment through numerical simulation (Miao, H. 2018). There is still room for improvement in warning accuracy and efficiency. Currently, numerous successful cases demonstrate the application of artificial intelligence to practical problems, such as subway passenger flow prediction and air pollution prediction. However, the use of artificial intelligence in wastewater treatment remains relatively limited. Therefore, it is worth considering combining artificial intelligence with wastewater treatment. Some researchers have also applied machine learning algorithms to wastewater treatment warnings. Hilal Anwer Mustafa et al. proposed a fusion model combining B-KNN and Extreme Learning Machine (ELM) algorithm, which showed the highest prediction accuracy of 93.56% in water quality classification. (Gao, Y., & Wang, J. 2025). Liu Ze Jun et al. established a prediction model of effluent chemical oxygen demand (COD) of anaerobic wastewater treatment system by using least squares support vector machine (LS-SVM), which provided a meaningful reference for improving the monitoring level of anaerobic wastewater treatment process (Ze-Jun, L., & others. 2019). Both of these studies employed machine learning methods for predictions, but they focused on time series data. As time series data, certain deep learning models often perform better than machine learning models. Given these issues, this paper analyzes and predicts factory wastewater discharge data using an improved LSTM model based on the acquired dataset.

3. Model Introduction

3.1. LSTM

Traditional Recurrent Neural Networks (RNN) often encounter difficulties when processing long sequential data, with the most typical problems being gradient vanishing and explosion. To address this, researchers designed a special type of RNN, known as Long Short-Term Memory (LSTM). LSTM can effectively capture long-term dependencies within data. Through the introduction of special memory cells and gating mechanisms, LSTM can learn and retain information over long time intervals more accurately. This capability is vital for tasks such as sequence prediction and other time series analyses. LSTM was proposed by Hochreiter and

Schmidhuber in 1997 and has found extensive application in speech recognition, machine translation, time series forecasting, and more.

The core idea of LSTM introduces "cell state," which serves as a continuous line of state allowing information to flow nearly losslessly. To control how information flows into or out of the cell state, LSTM has developed a set of gating mechanisms: the input gate, forget gate, and output gate. Each gate functions as a Sigmoid layer to determine how much information passes through. The basic structure of LSTM consists of the forget gate, input gate, and output gate (see Figure 1). With these gating structures, LSTM can selectively retain or discard information from the previous time step and pass relevant information to subsequent steps. This way, LSTM can effectively manage the flow of information, deciding what should be stored for the long term, what should be updated, or what should be ignored, thereby optimizing the handling of long sequential data.

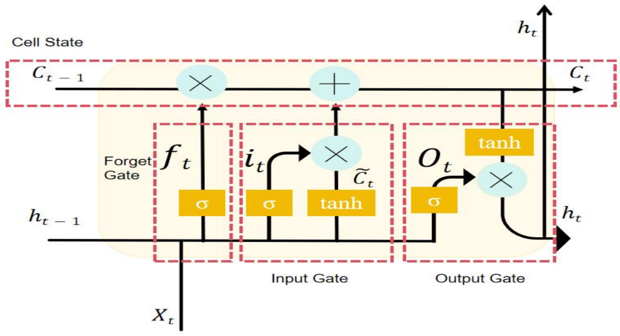

Figure 1. LSTM cell structure

3.2. Transformer

The Transformer model represents a significant breakthrough in deep learning. It discards traditional recursive neural networks (RNN) and convolutional neural networks (CNN) in favor of a self-attention-based architecture, greatly enhancing parallelization and the ability to handle long-range dependencies (Yuan, F., & Wang, Y. (2025). The Transformer model consists of an input part, an encoder-decoder structure, and an output part, with the encoder-decoder being the most critical section. The encoding part processes the input sequence and generates a context-aware representation. Each encoder consists of multiple identical layers stacked, with each layer containing two sublayers: a Multi-Head Self-Attention mechanism and a Feed-Forward Neural Network.

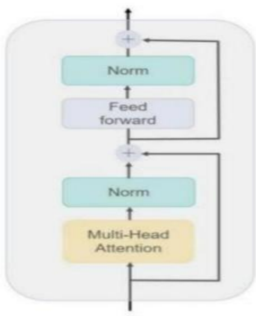

Figure 2. Encoder module structure

The Multi-Head Attention mechanism aggregates information by mapping the given query matrix $Q \in R^{d_q}$,key matrix $K \in R^{d_k}$, and value matrix $V \in R^{d_v}$ to the output.

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

When the query matrix Q, key matrix K, and value matrix V are equal, we refer to this as the attention mechanism. The Multi-Head Attention mechanism aggregates information by applying different linear transformations to map the queries, keys, and values into multiple independent subspaces. Each "head" learns specific representations within its own subspace, capturing different types of associations. Subsequently, these transformed queries, keys, and values are concatenated and integrated through an additional learnable linear transformation to produce the final output.

$$z_i = \text{Attention}\left(QW_i^Q, KW_i^K, VW_i^V\right)$$

$$Z = \text{Concat}\left(z_1, z_2, \ldots z_i\right)W^O$$

where $W_i^Q \in R^{p_q \times d_q}, W_i^K \in R^{p_k \times d_k}, W_i^V \in R^{p_v \times d_v}$ are the parameter matrices of query, key and value linear transformations respectively; Z_i represents the I-th attention head; ; Concat(.) represents an aggregation function for each attention head; $W^O \in R^{p_o \times kp_v}$ is the parameter matrix of the final linear transformation of the multi-head attention mechanism.

3.3. LSTM-Transformer mode

For traditional Transformer models, due to architectural characteristics, it is often challenging to capture temporal dependencies. As a result, data typically undergoes position encoding before being fed into the Encoder layers. To address this issue, this paper presents a hybrid predictive model that combines LSTM and Transformer. For data with significant time series characteristics, such as hourly flow and ammonia nitrogen emissions, an initial processing is carried out using the LSTM network. The LSTM can effectively capture long-term dependencies within time series data. Given that traditional Transformer models struggle to directly capture temporal dependencies in the input data, prior to passing the data to the Encoder layers of the Transformer, a specifically designed position encoding layer is necessary to capture time information within the time series. This step is crucial as it allows the Transformer to understand the temporal sequence in the input data. The data processed by the LSTM is then entered into the Encoder section of the Transformer model, which is composed of several identical layers stacked, each of which includes a Multi-Head Self-Attention mechanism and a Feed-Forward Neural Network. During this process, each "head" learns specific representations within its own subspace, thereby capturing relationships among different types of data, which can be enhanced through the Multi-Head Self-Attention mechanism. Finally, the features obtained after processing through the Transformer

Encoder layers are passed to the linear layer to generate the final prediction result (as shown in Figure 3)

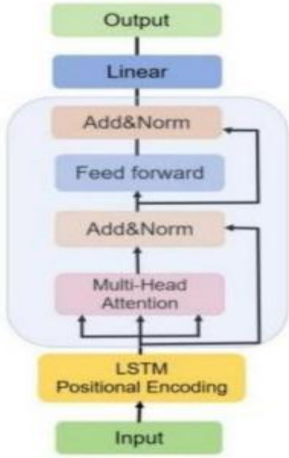

Figure 3. Overall structure of the model

4. Experiments and Results Analysise

4.1. Data Description and Presentation

The dataset used in this paper's experiments is a publicly available dataset on factory wastewater discharge provided by the Mathematical Modeling Competition of China University of Petroleum. This paper conducts an in-depth analysis of six different types of data, including hourly flow (time-flow), pH value (PH_value), ammonia nitrogen concentration (NHNI), ammonia nitrogen emissions (NHNIEmissions), chemical oxygen demand concentration (CODC), and chemical oxygen demand emissions (CODCEmissions). Data for hourly flow and pH value within a 24-hour period are selected, and visualization techniques are employed to present the trends of all data in the form of charts (as shown in Figure 4).

Figure 4. Characteristics of factory pollutant discharge data

4.2. Results Analysis

This paper uses the LSTM-Transformer model to predict various types of data in the dataset, starting with standardization of the data. We select feature columns (hourly flow, pH value, ammonia nitrogen concentration, ammonia nitrogen discharge, chemical oxygen demand concentration, and chemical oxygen demand discharge) and calculate the mean and standard deviation for each column. Using these two statistics, the data is transformed into a form with zero mean and unit variance. The dataset is then split into training and testing sets in a 7:3 ratio for training and testing purposes.

After completing the above data processing, a deep learning model combining LSTM and Transformer is constructed. This model consists of seven hidden layers and two output layers, utilizing Sigmoid and Tanh activation functions. The Adam optimizer is employed, which is an adaptive learning rate optimization algorithm that dynamically adjusts the learning rate during training to accelerate the model's convergence speed. The model is set for 100 epochs with a batch size of 32, and the testing set is used for model evaluation to assess the model's accuracy on the testing set. The model is used to make predictions on the testing set, and the actual and predicted values are output to a CSV file named contaminate.csv for subsequent calculation of evaluation metrics.

Table 1. Experimental configuration table

Configuration Item	Configuration Name
Operating System	Windows11
Processor	13th I7-13620H
Graphics Card	RTX4060
Internal Memory	16GB
Programming Language	Python 3.12
Development Environment	PyCharm

The relevant hardware and software configurations used in this experiment are shown in Tableabove. The experiment selected Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared (R^2) as evaluation metrics. These metrics measure the differences between predicted and actual values and reflect the accuracy of the model to varying degrees. To verify the effectiveness of the proposed LSTM-Transformer combined prediction model, comparative experiments were conducted using both the LSTM-Transformer model and the traditional Transformer model. The experimental results are shown in Table 2.

Table 2 Comparison of Evaluation Metrics for Different Models

Test Index	**MAE**	**RMSE**	**R^2**
		Transformer	
time_flow	0.283601	0.362260	0.832510
PH_value	0.265453	0.645592	0.737542
NHNI	0.105380	0.417337	0.694254

CODC	0.088356	0.264517	0.908036
NHNI Emissions	0.183624	0.452822	0.675283
CODC Emissions	0.206502	0.323472	0.865023
	LSTM-Transformer		
time_flow	0.254369	0.334975	0.875155
PH_value	0.249030	0.624723	0.765421
NHNI	0.073501	0.391269	0.730892
CODC	0.076008	0.238104	0.920108
NHNI Emissions	0.169790	0.426726	0.696941
CODC Emissions	0.178438	0.292899	0.903018

Based on the comparison of the prediction results of the two models shown in Table 2, it can be seen that the LSTM-Transformer model proposed in this paper has lower errors in predicting six types of data compared to the control group. Therefore, this model has a certain level of reliability.

5. Conclusion

This article focuses on predicting factory wastewater emissions. To more efficiently forecast the flow of factory discharge data, the paper proposes a combined prediction model based on LSTM-Transformer by deeply mining the multidimensional data within the dataset. This model can effectively capture local dynamic characteristics and global contextual dependencies in time series data. By introducing positional encoding and a multi-head attention mechanism, the model enhances the understanding of time series features, supports parallel computing, improves training efficiency, and mitigates the gradient vanishing problem in deep networks. The use of regularization techniques such as dropout ensures the model's generalization ability, allowing it to perform well on unseen data. Experimental results indicate that with appropriate hyperparameter tuning, the LSTM-Transformer model can provide stable prediction performance while maintaining high accuracy, as demonstrated in the prediction of factory wastewater emission data.

In the future, further research can explore integrating external environmental factors to improve prediction robustness. Enhancing model interpretability will be crucial for practical deployment in industrial settings. Additionally, real-time deployment of the LSTM-Transformer model on edge devices presents a promising direction for intelligent wastewater management.

Author Contributions:

Conceptualization, H. Z., F. Y.; methodology, H. Z., F. Y.; software, H. Z., F. Y.; validation, H. Z., F. Y.; formal analysis, H. Z., F. Y.; investigation, H. Z., F. Y.; resources, H. Z., F. Y.; data curation, H. Z., F. Y.; writing—original draft preparation, H. Z., F. Y.; writing—review and editing, H. Z., F. Y.; visualization, H. Z., F. Y.; supervision, H. Z., F. Y.; project administration,

H. Z., F. Y.; funding acquisition, H. Z., F. Y. All authors have read and agreed to the published version of the manuscript.

Funding:

This research was funded by 2024 Taiyuan Normal University Graduate Education Innovation Project (SYYJSYC-2475).

Institutional Review Board Statement:

Not applicable.

Informed Consent Statement:

Not applicable.

Data Availability Statement:

This dataset comes from the "2024 Asia-Pacific Cup Chinese Competition Mathematical Modeling B (Data Analysis and Prediction of Flood Disasters)".

Conflict of Interest:

The authors declare no conflict of interest.

References

Baosong, Y., Longyue, W., F. D. W., & others. (2021). Context-aware self-attention networks for natural language processing. Neurocomputing, 458, 157–169.

David, D., N. A. F., Jane, H., & others. (2016). CellProfiler Analyst: Interactive data exploration, analysis and classification of large biological image sets. Bioinformatics, 32(20), 3210–3212.

Gao, Y., Jiang, Y., Peng, Y., Yuan, F., Zhang, X., & Wang, J. (2025). Medical Image Segmentation: A Comprehensive Review of Deep Learning-Based Methods. Tomography, 11(5), 52.

Hui, L., Shigan, Y., & Linguo, L. (2018). Sewage particle concentration measurement based on image processing technology. Chemical Engineering Transactions, 71, 883–888.

Joseph, G., Moyez, D., Paulo, N., & others. (2014). Exploring high dimensional data with Butterfly: A novel classification algorithm based on discrete dynamical systems. Bioinformatics, 30(5), 712–718.

LiMin, Z., Yang, L., YueTing, Z., & others. (2023). A deep learning method using gender-specific features for emotion recognition. Sensors, 23(3), 1355.

Mahjouri, M., Ishak, B. M., Torabian, A., & others. (2017). The application of a hybrid model for identifying and ranking indicators for assessing the sustainability of wastewater treatment systems. Sustainable Production and Consumption, 10, 21–37.

Miao, H. (2018). Coal chemical wastewater treatment process based on computer simulation technology. Chemical Engineering Transactions, 67, 469–474.

Stefano, P., Giacomo, P., & Alessandro, P. (2021). LoPATraN: Low power asset tracking by means of Narrow Band IoT (NB-IoT) technology. Sensors, 21(11), 3772.

Yuan, F., Huang, X., Zheng, L., & Wang, Y. (2025). The evolution and optimization strategies of a PBFT consensus algorithm for consortium blockchains. Information, 16, 268.

Yuan, F., Zuo, Z., Jiang, Y., Shu, W., Tian, Z., Ye, C., Yang, J., Mao, Z., Huang, X., Gu, S., & Peng, Y. (2025). AI-Driven Optimization of Blockchain Scalability, Security, and Privacy Protection. Algorithms, 18(5), 263.

Ze-Jun, L., Jin-Quan, W., Yong-Wen, M., & others. (2019). Online prediction of effluent COD in the anaerobic wastewater treatment system based on PCA-LSSVM algorithm. Environmental Science and Pollution Research, 26(13), 12828–12841.