Journal of Information & Systems Management

DLINE JOURNALS

ISSN: 2230 – 8776

JISM 2025: 15 (4)
https://doi.org/10.6025/jism/2025/15/4/153-160

Liquid Quality Prediction Using Gated Recurrent Unit (GRU) Networks: A Comparative Study with LSTM and RNN Models

Wei Xiao
College of Physical Education, Fuyang normal University
Fuyang, Anhui. China
XBW289381025@163.com

ABSTRACT

The paper investigates water quality prediction using machine learning, with a focus on the Gated Recurrent Unit (GRU) model. As water pollution intensifies due to industrialization and urbanization, accurate forecasting becomes critical for environmental protection and sustainable water resource management. The study reviews existing approaches, including mechanistic models like Streeter Phelps and WASP, and data driven non mechanistic models such as ARIMA, SVR, and deep learning methods. Among deep learning architectures, GRU a simplified variant of LSTM offers strong performance in capturing long term temporal dependencies in time series data while maintaining lower computational complexity. The authors construct a GRU based water quality prediction model using monthly pH and dissolved oxygen (DO) data from China's Surface Water Quality Automatic Monitoring system (2011–2018). After preprocessing and normalization, the GRU model is trained on data from 2011–2017 and tested on 2018 data. Results show that GRU outperforms traditional RNNs and LSTMs in prediction accuracy for both pH and DO levels. The study concludes that GRU provides an efficient and precise solution for water quality forecasting, thereby supporting better decision-making in environmental management. Future work may focus on optimizing model architecture and feature selection to enhance robustness and generalization further. This research highlights the potential of deep learning, particularly GRU networks, in tackling real world environmental challenges.

Keywords: Water Quality Prediction, Machine Learning, GRU Model, Time Series Analysis Dissolved Oxygen, pH Levels, Environmental Protection, Deep Learning

Received:5 August 2025, Revised 28 September 2025, Accepted: 6 October 2025

1. Introduction

As industrialization and urbanization continue to progress rapidly, issues related to water quality have garnered significant global attention. Water is a vital resource for sustaining the earth's ecological balance and human health; however, its pollution is becoming increasingly severe. Predicting water quality, a crucial component of managing and safeguarding the water environment, is vital to ensuring the sustainable use of water resources and maintaining a healthy living environment. With the rapid advancements in computer science and artificial intelligence, machine learning algorithms have demonstrated considerable potential and application prospects for water quality prediction [1]. This paper intends to investigate and analyze water quality prediction methods utilizing machine learning algorithms, providing a scientific foundation and technical assistance for effectively addressing water environment challenges. Specifically, we will focus on the feasibility and efficacy of machine learning for water quality prediction, selecting the GRU (Gated Recurrent Unit) model as the primary object of study. As a significant variant of recurrent neural networks, GRU incorporates memory units and gating mechanisms, enabling it to effectively capture long term dependencies in time series data and excel at time series prediction [2]. Through thorough research on the GRU model, we aim to provide an efficient and precise solution for water quality prediction.

The importance of this research lies in the utilization of machine learning algorithms, which can enhance our understanding of the patterns and trends in water quality prediction, offering a scientific decision making foundation for water environment management authorities, optimizing the distribution and use of water resources, thus mitigating the risk of water environmental pollution and enhancing human living conditions [3]. Moreover, this study will examine the performance of the GRU model for water quality prediction and compare it with other widely used machine learning algorithms, thereby serving as a reference for research in this domain. In this paper, we will use substantial water quality monitoring data to evaluate the GRU model's effectiveness in predicting water quality. Simultaneously, we will conduct comparative experiments with other machine learning algorithms, such as LSTM (Long Short-Term Memory) and traditional time series prediction techniques, to comprehensively evaluate the performance and advantages of the GRU model. We believe that by exploring this research in depth, we will contribute new ideas and methodologies to advance water quality prediction.

2. Status of Water Quality Prediction Models based on Deep Learning

In recent years, water quality prediction models based on deep learning have emerged as a major research focus in the field of water environment. The implementation of deep learning technology has significantly enhanced water quality prediction capabilities. Such models have been extensively studied and attracted attention both nationally and internationally. Currently, water quality prediction models can be categorized into two main types: mechanistic models and non mechanistic models. Mechanistic models are relatively complex and require the development of mathematical models that account for the internal and external conditions of water bodies.

These models are utilized to investigate the connections between shifts in water quality and various influencing factors, as well as to forecast the effects of environmental changes on aquatic systems. In the initial phases of research focused on predicting water quality, these models were frequently employed.

As early as 1925, H. Streeter and colleagues developed the first model of river water quality, the Streeter Phelphs model, which illustrated variations in biochemical oxygen demand and dissolved oxygen in one dimensional steady state rivers and achieved widespread use. Following this, countries in Europe and America progressively developed models such as COASTOX, SWAP, WASP, AQUATOX, among others. [4] For instance, the WASP model introduced by Di Toro and others is capable of constructing both linear and nonlinear dynamic models for investigating 1D, 2D, and 3D systems along with numerous types of pollutants, making it one of the most extensively utilized water quality models globally. Concurrently, the U. S. Environmental Protection Agency formulated the AQUATOX model, which can forecast variations in diverse pollutants and their effects on ecosystems. Nonetheless, the development of mechanistic models is intricate and demands a significant amount of specialized theoretical knowledge. [5] With advancements in machine learning and related technologies, the focus of water quality prediction research has shifted to non mechanistic models that rely on historical data. Researchers from various nations have incorporated techniques such as regression analysis, time series analysis, support vector regression, and artificial neural networks into the study of water quality prediction, considerably simplifying the modeling process. [6] These non mechanistic models can proficiently predict alterations in water quality by recognizing patterns and trends in historical data, thereby providing robust support for the management of water resources and environmental conservation.

Compared with other countries, China began its efforts in water quality prediction and warning relatively late, establishing systems for rivers such as the Yellow River and the Three Gorges only at the close of the 20th century. Since then, Chinese researchers have eagerly adopted and integrated numerous advanced technologies and insights, achieving notable advances in water quality prediction and alerting. Recently, as China's economy has rapidly advanced, pollution of surface water from industry and agriculture has also escalated significantly, posing greater challenges for water resource protection. Consequently, water quality prediction has garnered substantial attention from both the government and academia. [7] Among the frequently employed prediction techniques, time series analysis focuses on exploring relationships within data sequences. For instance, Shi Zibo and Zou Zhihong utilized wavelet transform to differentiate high frequency and low frequency components of dissolved oxygen data in the Heihe River Basin in Heilongjiang, and developed an ARIMA model to project future changes in dissolved oxygen concentration, achieving high predictive accuracy. [8] Guo Yang and Ding Zhenjun utilized historical monitoring data from various segments of the Liao River Basin to develop ARIMA forecasting models aimed at predicting future water quality conditions for key indicators such as ammonia nitrogen, chemical oxygen demand, biochemical oxygen demand, and total phosphorus. The findings indicated that the ARIMA model can forecast the concentrations of these indicators. Among the commonly employed algorithms, support vector regression (SVR), an extension of support vector machines (SVMs), has also found extensive application in water quality forecasting. Some researchers examined a specific station on the Haihe River, using ammonia nitrogen and total phosphorus concentrations as data to construct an SVR based water-quality forecasting model. The experimental results demonstrated that the SVR model's predictive performance accurately reflected water quality conditions. Furthermore, additional researchers focused on the critical parameters of dissolved oxygen and pH levels in aquaculture water as the target for prediction, employing ensemble empirical mode decomposition (E-EMD) to denoise the sample data, selecting SVR as the method for water quality forecasting, and utilizing the radial basis kernel function through grid search to determine the optimal parameters. The experimental results indicated that the combined EEMD-SVR approach was more effective for forecasting water quality indicators in aquaculture ponds, outperforming the SVR method without EEMD.

3. Water Quality Prediction Model Based on Machine Learning Algorithms

3.1 Theoretical Basis of GRU Network

While domestic research on deep learning for water quality forecasting began relatively recently, the ongoing advancements in environmental science and artificial intelligence technology in China suggest that domestic studies in this area will continue to achieve new breakthroughs.Concurrently, collaboration and communication between domestic and international scholars will further enhance the development of this field. The GRU (Gated Recurrent Unit) is an enhanced model of the Recurrent Neural Network (RNN), designed to address the challenges of gradient vanishing and long term dependencies encountered in conventional RNNs. The essence of the GRU network lies in its internal gating mechanisms, which effectively capture long term dependencies in time series data. [9] In the GRU framework, each time step has a hidden state that is updated based on the current input and the hidden state from the previous time step. The reset gate determines whether to retain information from the past, generating a value between 0 and 1 by processing the input and the hidden state from the previous time step, thereby indicating the weight of past information. The update gate, similarly derived from the input and the previous time step's hidden state, produces a value between 0 and 1 that signifies the weight of the candidate hidden state.

The candidate hidden state represents a newly calculated value derived from the current input and the hidden state from the preceding time step. This value is then blended with the hidden state from the previous time step, guided by the update gate weight, to produce the final hidden state for the current time step. This gating mechanism enables the GRU network to flexibly and selectively forget earlier information and to assess the influence of both past data and the current input at the current time step. Consequently, the GRU network excels in capturing long term dependencies within time series data, effectively addressing the challenges of gradient vanishing and long term dependencies that often hinder conventional RNNs. Compared with traditional RNNs and LSTM (Long Short Term Memory) networks, the GRU architecture is more streamlined, incurs lower training costs, and has reduced computational complexity, while delivering outstanding results across various sequence data tasks, as illustrated in Figure 1.

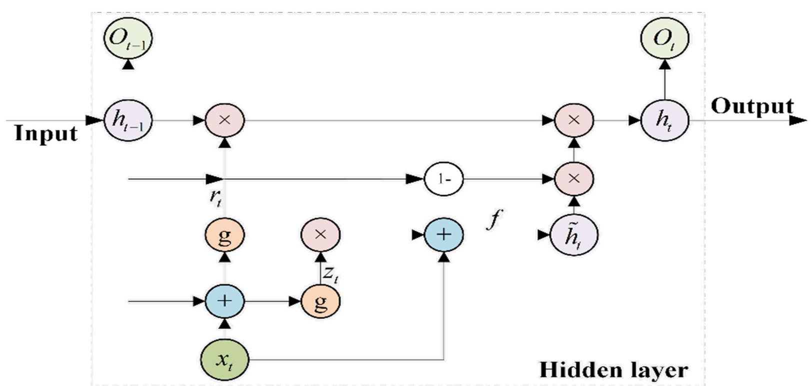

Figure 1. GRU model framework diagram

In Figure 1, the reset gate and update gate are and, respectively. The value of the output layer is, and the value of the hidden layer is. Subtracting 1 from each element in the vector is represented by the symbol "1-". The two vectors are represented by the symbol "" (matrix calculation method), and the sum operation of the two vectors is represented by "+".

$$g = sigmoid(x) = \frac{1}{1+e^{x}} \tag{1}$$

$$f = \tanh(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} \tag{2}$$

From the figure, it can be observed that the update gate parameters affect the incorporation of the previous time step's state information, with a proportional relationship. A larger update gate leads to a greater incorporation of information. The state information from the prior time step will be written into the candidate set, and the size of the reset gate also influences the writing of the last time step's state information; their relationship is a proportional one a larger reset gate results in more state information being written.

3.2 Water Quality Prediction Model based on GRU Algorithm

The GRU algorithm is a type of recurrent neural network primarily used for handling time series data. Its process is as follows: First, it takes the feature vector of the current time step and the hidden state from the previous time step as inputs. Then, the GRU adjusts information propagation and forgetting via gating mechanisms, including the update and reset gates. The update gate controls the degree to which the current time step's input features update the hidden state, while the reset gate determines whether to ignore the hidden state from the previous time step. Next, it computes the new hidden state for the current time step using the gating mechanisms. Finally, the new hidden state becomes the input for the next time step, and this iteration continues until the end of the sequence. The gating mechanisms of the GRU algorithm can better capture long term dependencies, enabling the model to effectively handle time series data, such as temperature, pH, and other indicators, for water quality prediction, thereby improving prediction accuracy and stability.

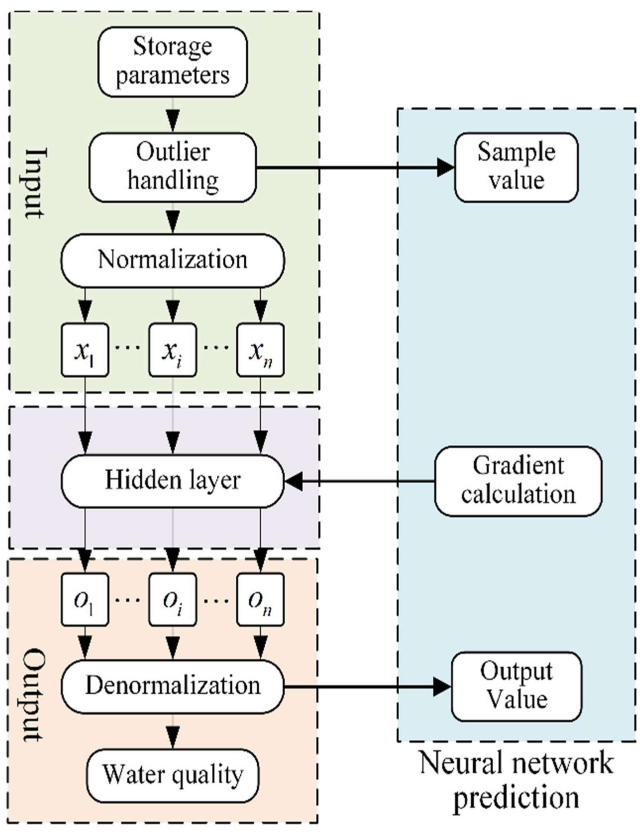

Figure 2. Water Quality Prediction Process based on GRU Algorithm

The water quality prediction process based on the GRU algorithm is illustrated in Figure 2. First, in the data preprocessing stage, collected water quality monitoring data undergoes outlier treatment, where statistical methods or machine learning techniques are used to detect potential abnormal data points, and appropriate strategies such as deletion, interpolation, or replacement are applied. Next, normalization is performed to scale the original water quality data features to a uniform range. Common methods include Min Max normalization or Z-score normalization. Then, based on the prediction target and data features, a GRU neural network model is constructed, specifying the configurations of the input, hidden, and output layers, as well as the activation functions. The model enters the training stage, where it is trained using the training dataset, and model parameters are continuously adjusted through optimization algorithms to enhance predictive performance. Once the training is complete, the water quality prediction is performed using the test dataset. The input data is fed into the trained GRU model to obtain predictions. Finally, reverse normalization is applied to restore the prediction results to the original data range, obtaining the final water quality prediction results. The prediction results are then evaluated by comparing the predicted values with the observed values, using appropriate metrics such as Root Mean Square Error (RMSE) or Mean Absolute Error (MAE) to assess the model's predictive performance. If the prediction results are satisfactory, the model can be used for water quality prediction tasks; otherwise, model structures or parameters can be adjusted for optimization until the desired prediction accuracy is achieved.

4. Comparative Analysis of Water Quality Prediction Experimental Results based on GRU Model

In our research on predicting water quality, we utilized water quality monitoring data sourced from the Surface Water Quality Automatic Monitoring and Publishing Website of the Environmental Monitoring Center under the Ministry of Ecology and Environment, covering the years 2011 to 2018 as our experimental dataset (refer to Figure 3). This dataset is organized by month and includes two key water quality metrics: pH and Dissolved Oxygen (DO). To assess the effectiveness of our proposed method, we also implemented LSTM and RNN models for comparative analysis. The experimental procedure unfolded as follows: Initially, we performed essential data preprocessing on the original dataset, including handling outliers and missing values to ensure data integrity and precision. Following this, we normalized the data to standardize the water quality metrics within a consistent range, enhancing their applicability in the GRU model. We then established a water quality prediction model based on the GRU algorithm. This model features an input layer, a hidden layer with GRU units to capture temporal dynamics, and an output layer for forecasting future values of the water quality metrics. During the training phase, data from 2011 to 2017 served as the training set to develop the GRU model, with model parameters being continuously refined through optimization algorithms to enhance predictive accuracy. Once training was complete, data from 2018 were utilized as the test set, and the trained GRU model was employed for water quality predictions. The results of these predictions emerged as normalized values. In the final step, reverse normalization was conducted to revert the prediction results to the original range of the water quality indicators, yielding the ultimate water quality prediction outcomes.

In our experiment, we used water quality data from all 12 months of 2018 as the test set to validate the model. The results indicate that the GRU model exhibits the highest accuracy in predicting pH and dissolved oxygen levels, demonstrating the closest alignment with actual values. Its performance significantly surpasses that of traditional RNN and LSTM models.

This experimental outcome illustrates the notable benefits of the GRU model for predicting water quality. The gating mechanism inherent in the GRU model is adept at capturing long term dependencies in time series data, and its relatively streamlined architecture minimizes training expenses and computational demands. These traits

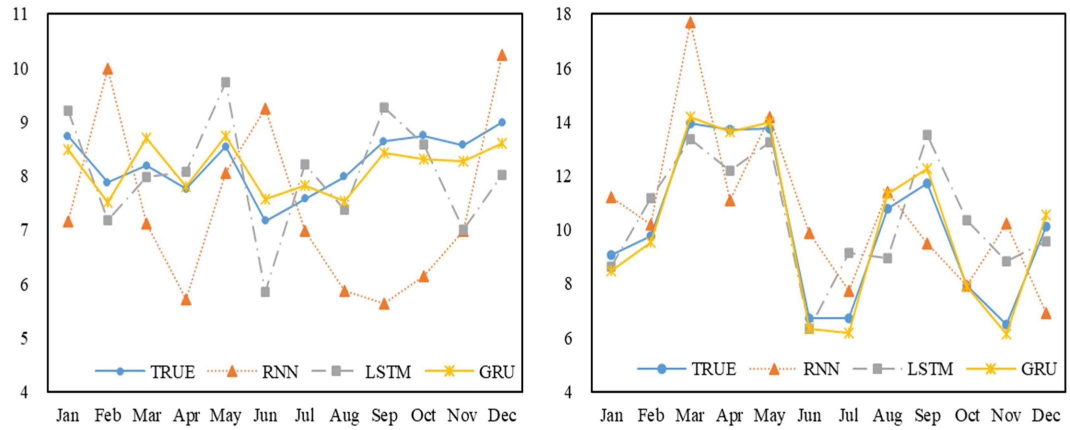

Figure 3. Comparison of Predicted Values and True Values using Different Methods (Left: pH Value, Right: Dissolved Oxygen Value / mg·L-1)

empower the GRU model to effectively learn trends in water quality data, resulting in more precise predictions. Particularly impressive is the GRU model's outstanding ability to forecast both pH levels and dissolved oxygen concentrations. These parameters are vital indicators of aquatic environment health, and accurate predictions are critical for managing water resources and conserving the environment. The GRU model's high accuracy on these two metrics provides a reliable tool for monitoring water quality and preventing and controlling pollution.

5. Conclusion

This research validated a water-quality prediction model based on the GRU algorithm using 2018 data and contrasted it with conventional RNN and LSTM models. The findings reveal that the GRU model achieves the highest accuracy in predicting pH levels and dissolved oxygen concentrations, aligning closely with the actual values and significantly surpassing the performance of RNN and LSTM models. The results clearly underscore the GRU model's dominance in water quality prediction, demonstrating an efficient and precise approach for these tasks. Nevertheless, it is essential to recognize that variables such as data quality and feature selection can still influence the prediction outcomes. Future investigations may focus on further optimizing the model's structure and parameters, as well as exploring more effective methods for feature selection to enhance the robustness and generalization capabilities of water quality prediction models. In conclusion, this study confirms the superiority of the GRU model in water quality prediction tasks, offering significant insights and practical applications for water resource management, environmental protection, water quality monitoring, and related areas. By integrating deep learning algorithms, we can more effectively harness water-quality data and make informed decisions to support the health of aquatic ecosystems and sustainable environmental development.

Acknowledgements

The study was supported by the Research on the application of rainwater collection and utilization system in the construction of water-saving community (No. KJQN202003809).

References

[1] Chen, Y., Song, L., Liu, Y., et al. (2020). A review of the artificial neural network models for water quality prediction. *Applied Sciences,* 10 (17), 5776.

[2] Fu, B., Merritt, W. S., Croke, B. F. W., et al. (2019). A review of catchment scale water quality and erosion models and a synthesis of future prospects. *Environmental modelling software,* 114, 75-97.

[3] Najafzadeh, M., Ghaemi, A. (2019). Prediction of the five day biochemical oxygen demand and chemical oxygen demand in natural streams using machine learning methods. *Environmental monitoring and assessment,* 191, 1-21.

[4] Chen, W., Chen, H. J., Dai, F. X. (2020). Effluent water quality prediction model based on artificial neural network for wastewater treatment. W*ater Wastewater Engineering,* 46, 990-994.

[5] Azrour, M., Mabrouki, J., Fattah, G., et al. (2022). Machine learning algorithms for efficient water quality prediction. *Modeling Earth Systems and Environment,* 8 (2), 2793-2801.

[6] Newhart, K. B., Torres, Goldman., J. E., Freedman, D. E., et al. (2020). Prediction of peracetic acid disinfection performance for secondary municipal wastewater treatment using artificial neural networks. *ACS EST Water,* 1 (2), 328-338.

[7] Chowdury, M. S. U., Emran, T. B., Ghosh, S., et al. (2019). IoT based real time river water quality monitoring system. *Procedia computer science,* 155, 161-168.

[8] Tang, T., Strokal. M., Vliet, van., M. T. H., et al. (2019). Bridging global, basin and local scale water quality modeling towards enhancing water quality management worldwide. *Current opinion in environmental sustainability,* 36, 39-48.

[9] Rahman, A., Dabrowski, J., McCulloch, J. (2020). Dissolved oxygen prediction in prawn ponds from a group of one step predictors. *Information Processing in Agriculture,* 7 (2), 307-317.

[10] Kurra, S. S., Naidu, S. G., Chowdala, S., et al. (2022). Water quality prediction using machine learning. Int. Res. J. Mod. Eng. Technol. *Sci,* 4, 692-696.