Deep-Learning-Based Water Quality Monitoring and Early Warning Methods: A Case Study of Ammonia Nitrogen Prediction in Rivers

Xianhe Wang 1,2, Mu Qiao 2, Ying Li 1,2, Adriano Tavares 2, Qian Qiao 1 and Yanchun Liang 3,4,*

1 School of Applied Chemistry and Materials, Zhuhai College of Science and Technology, Zhuhai 519041, China; wxh@z cst.edu.cn (X.W.)
2 Department of Industrial Electronics, School of Engineering, University of Minho, 4704-553 Braga, Portugal
3 School of Computer Science, Zhuhai College of Science and Technology, Zhuhai 519041, China
4 Key Laboratory of Symbol Computation and Knowledge Engineering of the Ministry of Education, College of Computer Science and Technology, Jilin University, 2699 Qianjin Street, Changchun 130012, China

* Correspondence: ycliang@jlu.edu.cn

Abstract: In line with rapid economic development and accelerated urbanization, the increasing discharge of wastewater and agricultural fertilizer usage has led to a gradual rise in ammonia nitrogen levels in rivers. High concentrations of ammonia nitrogen pose a significant challenge, causing eutrophication and adversely affecting the aquatic ecosystems and sustainable utilization of water resources. Traditional ammonia nitrogen detection methods suffer from limitations such as cumbersome sample handling and analysis, low sensitivity, and lack of real-time and dynamic feedback. In contrast, automated monitoring and ammonia nitrogen prediction technologies offer more efficient methods and accurate solutions. However, existing approaches still have some shortcomings, including sample processing complexity, interference issues, and the absence of real-time and dynamic information feedback. Consequently, deep learning techniques have emerged as promising methods to address these challenges. In this paper, we propose the application of a neural network model based on Long Short-Term Memory (LSTM) to analyze and model ammonia nitrogen monitoring data, enabling high-precision prediction of ammonia nitrogen indicators. Moreover, through correlation analysis between water quality parameters and ammonia nitrogen indicators, we identify a set of key feature indicators to enhance prediction efficiency and reduce costs. Experimental validation demonstrates the potential of our proposed approach to improve the accuracy, timeliness, and precision of ammonia nitrogen monitoring and prediction, which could provide support for environmental management and water resource governance.

Keywords: artificial intelligence; LSTM model; neural networks; deep learning; applications of computational intelligence

1. Introduction

In recent years, rapid economic development and accelerated urbanization have led to improvements in industrial and agricultural production, as well as the living standards of urban residents. However, this progress has resulted in increased wastewater discharge and agricultural fertilizer usage, leading to a gradual rise in the concentration of ammonia nitrogen in rivers [1]. While ammonia nitrogen is an essential nutrient in river water, excessive levels can cause environmental issues, with water eutrophication being one of the most serious problems [2]. Eutrophication refers to the excessive nutrient content in river water, which triggers a rapid increase in biomass and fundamental changes in the aquatic ecosystem [3]. High concentrations of ammonia nitrogen promote the growth of algae and other aquatic plants, leading to an abundance of algae and phytoplankton, discoloration, and the emergence of harmful algae such as “Blue-Green Algae”.

Electronic version of this article is freely accessible at https://www.mdpi.com/journal/electronics

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland.
This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).
proliferation and death of these organisms result in a sharp decline in dissolved oxygen, deteriorating water quality and creating “Dead Zones”. These not only affect the river’s aquatic ecosystem but also have significant negative consequences for water resource utilization and ecological conservation. Moreover, excessive ammonia nitrogen levels pose risks to other organisms, including fish and invertebrates, affecting their respiratory and reproductive systems, and potentially causing respiratory difficulties, toxin accumulation, and even death. Additionally, ammonia nitrogen can react with other substances in water to form compounds such as nitrites and nitrates, which can harm human and animal health [4–6].

As a result, monitoring ammonia nitrogen concentrations in rivers has become a crucial task for environmental management. By monitoring ammonia nitrogen levels, pollution in river water can be promptly detected, enabling appropriate measures to be taken to prevent water eutrophication and other environmental problems. Furthermore, monitoring ammonia nitrogen levels provides scientific evidence for environmental management and protection, serving as a basis for formulating environmental protection policies and supporting sustainable water resource utilization [7–9].

To address water quality concerns, various water quality monitoring technologies, including ammonia nitrogen detection and early warning techniques, have been developed [10–12]. Traditional methods for ammonia nitrogen detection, such as the Nessler method, evaporation determination method, indicator method, and fluorescence method, have limitations in terms of cumbersome operations, low sensitivity, and limited accuracy. In recent years, automated monitoring technologies such as chromatography, electrochemical methods, optical methods, and biosensors have been widely adopted for ammonia nitrogen detection [13,14]. These methods offer advantages such as simplified operations, high efficiency, and improved accuracy, some of which enable real-time monitoring of water quality. Additionally, current ammonia nitrogen early warning technologies utilize a combination of monitoring instruments and information systems to achieve real-time monitoring and early warning of water quality conditions through data collection, transmission, processing, and analysis. Despite the numerous studies conducted on surface water ammonia nitrogen monitoring and early warning, practical applications still face limitations. Traditional chemical analysis methods involve laborious sample handling and analysis procedures, leading to potential errors. Novel techniques such as biosensors exhibit high sensitivity but encounter interference issues in complex environments. Furthermore, conventional monitoring methods often provide static data information and lack real-time and dynamic information feedback [15]. Therefore, improving the accuracy, timeliness, and precision of ammonia nitrogen monitoring and early warning in surface water remains an important research direction [16–18].

With the rapid development of artificial intelligence, machine learning has emerged as a popular technology in environmental and water resource management. Traditional machine learning methods have many advantages, such as ease of understanding and interpretation, visual analysis, and easy extraction of rules. In a relatively short period of time, these methods can produce feasible and effective results on large data sources and can handle both categorical and numerical data. They are suitable for handling missing samples and have a fast running speed when testing the dataset. However, there are also obvious disadvantages to machine learning, such as difficulties in handling missing data, the tendency to overfit, and ignoring the correlation between attributes in the dataset. Practical applications have shown that deep learning outperforms traditional machine learning and statistical methods in many tasks [19]. For example, deep learning models can learn and capture complex features of data, including nonlinear relationships and high-order interactions, which provides deep learning with greater flexibility and an advantage in dealing with complex, dynamic, and unknown data. It has strong representational power, and is able to handle data with high-dimensional features, nonlinear relationships, and complex patterns. It also has a high tolerance for noise and outliers, better adaptability to real-world applications, and improved robustness and generalization capabilities. As the amount of
data increases, traditional machine learning methods may encounter problems such as the curse of dimensionality. However, deep learning models have excellent scalability and can easily handle large-scale datasets, allowing them to learn more complex patterns from a large amount of data. It has strong memory capabilities, and is able to store and recall a large amount of information. This provides deep learning with a great advantage in application scenarios that require long-term memory and historical information. Finally, in many application scenarios, deep learning can achieve higher prediction accuracy than traditional machine learning methods; in particular, in the field of water quality prediction, deep learning algorithms perform significantly better than traditional machine learning algorithms [20]. In this study, a deep learning model called Long Short-Term Memory (LSTM) was employed to process water quality monitoring data and achieved high-precision prediction of ammonia nitrogen indicators through data analysis and modeling [21–23]. LSTM is a recursive neural network (RNN) that solves the problem of gradient disappearance or explosion that exists in traditional RNNs when dealing with long-sequence data by introducing memory units, allowing better capture of the time-series characteristics of the data when dealing with long-sequence data. At the same time, the gate control mechanism in the LSTM model can effectively control the flow of information, avoiding gradient disappearance or explosion problems. Therefore, when dealing with water quality data, LSTM can better capture the long-term dependence relationship between water quality indicators and improve predictive performance through forward and reverse information flow, thus more accurately predicting water quality data [24–26]. Furthermore, correlation analysis between different water quality indicators and ammonia nitrogen indicators helps in the identification of key feature indicators for model input, enhancing prediction efficiency and reducing costs [12,22]. To achieve these objectives, a series of experiments were conducted using historical monitoring data from the Qianshan River in Zhuhai City.

2. Materials and Methods
2.1. Study Area and Data Collection

The Qianshan River waterway plays a vital role as a major inland transportation route in Zhuhai City, China. It is located at 21°48’~22°27’ north latitude and 113°03’~114°19’ east longitude, in the south of Guangdong Province, on the west bank of the Pearl River estuary. Its source can be traced back to Lianshiwan in Tantou Town, Zhongshan City, where water is introduced from the Madaomen waterway and flows eastward, passing through Tantou Town and Qianshan Street in Zhuhai City, until it merges into the Pearl River Estuary at Wanzai Shikaoju lock. With a total length of approximately 23 km, the river encompasses a stretch of about 15 km in Tantou Town, Zhongshan, and varies in width from 58 to 220 m. In Zhuhai, the river extends for about 8 km with a width ranging from 200 to 300 m. The Qianshan River basin covers a watershed area of around 338 km², experiencing an annual runoff volume of 1.54 billion cubic meters, an average annual runoff depth of 1100 mm, and an average runoff coefficient of 0.58. The river basin predominantly consists of sedimentary plain landforms, sloping from the northeast to the southwest.

Since 2015, the Qianshan River basin has experienced a total of 107 industrial pollution sources. Out of these, 20 are located in Sanxiang Town, Zhongshan City, representing 18.7% of the total sources, while 45 are situated in Tantou Town, accounting for 42.1%. Additionally, the Zhuhai area hosts 42 industrial pollution sources, making up 39.3% of the overall count. Urban domestic pollution primarily consists of sewage from urban villages and scattered old villages along the river. Figure 1 shows the specific locations of monitoring areas and monitoring stations.

For the purpose of this study, water quality data was collected from the Shijiaoju monitoring point within the Qianshan Street waterway network. The dataset spans from 8 November 2020 to 28 February 2023, providing historical water quality data at four-hour intervals. The dataset comprises a total of 5058 samples, encompassing nine water quality parameters: ammonia nitrogen (NH₃-N), water temperature (Temp), potential of hydrogen
(pH), dissolved oxygen (DO), potassium permanganate index (KMnO₄), total phosphorus (TP), total nitrogen (TN), conductivity (Cond), and turbidity (Turb).

**Figure 1.** Spatial distribution of monitoring points within the study area.

2.2. Data Preprocessing

During the operation of automated water quality monitoring stations, various factors, including sensor malfunctions, network failures, and unexpected events such as pollutant leaks or extreme weather conditions, can lead to data loss and anomalies. The objective of data preprocessing is to cleanse the raw data by eliminating outliers, noise, and missing values, thereby improving the performance and reliability of water quality prediction models. Thorough data preprocessing ensures that the models are built upon high-quality data, enhancing prediction accuracy and providing a more dependable scientific foundation for water quality monitoring and management decisions [27–29].

In the context of handling missing values, two primary approaches, namely single imputation (SI) and multiple imputation (MI), are commonly used. While MI is more complex in operation and relatively costly, this study, considering the nature of the Qianshan River water quality data, adopts linear interpolation as the method for filling missing values. Linear interpolation, widely employed for filling missing values, is particularly suitable for data with a time dimension, such as time series data. Its fundamental concept involves estimating the missing values by performing linear interpolation between the preceding and subsequent observed values [30,31].

To implement linear interpolation, the positions of the missing values within the time series, referred to as interpolation positions, must be determined. Subsequently, the interpolation values are calculated by applying linear interpolation based on the available observed values, thereby obtaining estimates for the missing values [26,27]. Finally, it is essential to verify the interpolation results by ensuring that the post-interpolation data align with the actual situation, adhere to data distribution characteristics, and maintain consistency with other variables.

Let \((X₁, Y₁)\) represent the preceding observed value of the missing value, \((X₂, Y₂)\) represent the subsequent observed value, and \(X₀\) represent the position of the missing value. The estimated missing value \(Y₀\) can be calculated using the following formula:

\[
Y₀ = Y₁ + (X₀ − X₁) \times \frac{(Y₂ − Y₁)}{(X₂ − X₁)}
\]
Here, \( Y_1 \) and \( Y_2 \) represent the values of the observed values preceding and following the missing value, respectively, while \( X_1 \) and \( X_2 \) represent the corresponding time or position information. \( X_0 \) represents the position of the missing value [32]. Figure 2 and Table 1 show the basic situation of the water quality data.

![Temporal variation curves of water quality parameters.](image)

**Figure 2.** Temporal variation curves of water quality parameters.

**Table 1.** Summary statistics of water quality parameters.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Units</th>
<th>Max</th>
<th>Min</th>
<th>Mean</th>
<th>Std</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH(_3)-N</td>
<td>mg/L</td>
<td>3.095</td>
<td>0.025</td>
<td>0.462</td>
<td>0.462</td>
</tr>
<tr>
<td>Temp</td>
<td>°C</td>
<td>32.630</td>
<td>13.400</td>
<td>24.617</td>
<td>4.754</td>
</tr>
<tr>
<td>DO</td>
<td>mg/L</td>
<td>28.241</td>
<td>1.010</td>
<td>8.254</td>
<td>3.889</td>
</tr>
<tr>
<td>KMnO(_4)</td>
<td>mg/L</td>
<td>10.470</td>
<td>1.120</td>
<td>3.781</td>
<td>1.626</td>
</tr>
<tr>
<td>TP</td>
<td>mg/L</td>
<td>0.319</td>
<td>0.031</td>
<td>0.111</td>
<td>0.050</td>
</tr>
<tr>
<td>TN</td>
<td>mg/L</td>
<td>5.755</td>
<td>1.945</td>
<td>3.276</td>
<td>0.685</td>
</tr>
<tr>
<td>Cond</td>
<td>µs/cm</td>
<td>2988.800</td>
<td>214.301</td>
<td>1261.352</td>
<td>977.574</td>
</tr>
<tr>
<td>Turb</td>
<td>NTU</td>
<td>268.377</td>
<td>5.302</td>
<td>53.190</td>
<td>34.524</td>
</tr>
</tbody>
</table>

2.3. Feature Dataset

The dataset was thoroughly analyzed prior to model construction to gain insights into the relationships among variables, particularly focusing on the correlations between the input variables and the output variable [33]. Strong correlations between input and output variables indicate that the input values can effectively predict the output values, enabling the model to utilize this information during the modeling process [34]. Consequently, the model is expected to exhibit superior predictive performance by accurately capturing the relationships between inputs and outputs [22]. Conversely, weak correlations between
input and output variables imply limited predictive capability of the input variables for the output variables [35]. In such cases, the model may struggle to capture these relationships, resulting in restricted predictive performance as it fails to extract sufficient information from the input variables to accurately predict the output variables.

In this paper, the Pearson correlation coefficient, a widely used measure for assessing linear correlations between random variables, was employed to analyze the correlations among input variables and between input variables and the output variables. By calculating the Pearson correlation coefficient, we were able to evaluate the strength of correlations among input variables and the association between the input variables and the output variable [36]. This data analysis facilitated the identification of strong correlations among input variables, addressing the issue of redundant information and enhancing the model’s efficiency [23]. Table 2 shows the calculation results of Pearson correlation coefficient.

<table>
<thead>
<tr>
<th>Variables</th>
<th>NH$_3$-N</th>
<th>Temp</th>
<th>pH</th>
<th>DO</th>
<th>KMnO$_4$</th>
<th>TP</th>
<th>TN</th>
<th>Cond</th>
<th>Turb</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH$_3$-N</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temp</td>
<td>0.026</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pH</td>
<td>-0.420</td>
<td></td>
<td>-0.331</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DO</td>
<td>-0.394</td>
<td></td>
<td>-0.568</td>
<td>0.790</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KMnO$_4$</td>
<td>0.209</td>
<td>-0.326</td>
<td>0.527</td>
<td>0.511</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TP</td>
<td>0.613</td>
<td>0.426</td>
<td>0.409</td>
<td>-0.485</td>
<td>0.140</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TN</td>
<td>0.447</td>
<td>-0.199</td>
<td>0.199</td>
<td>0.057</td>
<td>0.547</td>
<td>0.466</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cond</td>
<td>-0.038</td>
<td>-0.617</td>
<td>0.627</td>
<td>0.628</td>
<td>0.760</td>
<td>-0.243</td>
<td>0.420</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Turb</td>
<td>-0.022</td>
<td>0.362</td>
<td>-0.345</td>
<td>-0.354</td>
<td>-0.203</td>
<td>0.333</td>
<td>-0.149</td>
<td>-0.418</td>
<td>1</td>
</tr>
</tbody>
</table>

The Table 2 analysis revealed significant correlations between NH$_3$-N and six parameters, namely pH, DO, KMnO$_4$, TP, TN, and Cond. Specifically, the correlation coefficient between NH$_3$-N and pH was -0.420, demonstrating a significant negative correlation ($p < 0.01$) between NH$_3$-N and pH. Similarly, NH$_3$-N and DO exhibited a correlation coefficient of -0.394, indicating a significant negative correlation ($p < 0.01$) between NH$_3$-N and DO. In contrast, NH$_3$-N and KMnO$_4$ showed a correlation coefficient of 0.209, suggesting a significant positive correlation ($p < 0.01$) between NH$_3$-N and KMnO$_4$. The correlation coefficient between NH$_3$-N and TP was 0.613, indicating a significant positive correlation ($p < 0.01$) between NH$_3$-N and TP. Moreover, NH$_3$-N and TN had a correlation coefficient of 0.447, signifying a significant positive correlation ($p < 0.01$) between NH$_3$-N and TN. Lastly, NH$_3$-N and Cond exhibited a correlation coefficient of -0.038, indicating a significant negative correlation ($p < 0.01$) between NH$_3$-N and Cond.

Conversely, no significant correlations ($p > 0.05$) were observed between NH$_3$-N and Temp or Turb, suggesting no significant relationship between NH$_3$-N and these two parameters.

Based on the results of the correlation analysis, each parameter was ranked according to the magnitude of their correlation coefficients. The parameters were then divided into nine groups, with increasing correlation coefficient values, as visually depicted in Figure 3. This grouping allows for a better understanding of the relationships between NH$_3$-N and other parameters, with parameters exhibiting higher correlation coefficients being considered more strongly associated with NH$_3$-N levels.
The design and training stages of deep learning models are pivotal in water quality modeling and prediction. Given the multifaceted influences and the temporal-spatial patterns inherent in NH$_3$-N concentrations in surface water, the adoption of the Long Short-Term Memory (LSTM) model, a prominent type of recurrent neural network (RNN), is judicious. Notably, LSTM boasts memory prowess, facilitating adept capture of long-term dependencies inherent in time series data [37]. Especially in the field of water quality prediction, the LSTM algorithm represents a significant improvement compared to traditional machine learning algorithms [38].

During the model training phase, historical NH$_3$-N monitoring data necessitate partitioning into training, validation, and testing sets, designated for model training, validation, and testing, respectively. This partitioning can be realized through either time-series-based or random division, ensuring that the data in these subsets remain representative both temporally and spatially. In this work, the validation set encompassed 10% of the dataset, totaling 506 samples, while the testing set comprised 5% of the dataset, amounting to 253 samples. The remaining samples were allocated for model training.

For model construction, training, and optimization, renowned deep learning frameworks such as TensorFlow and Keras come to the fore, streamlining efficient model design and training. Techniques including grid search and cross-validation prove instrumental in hyperparameter tuning. Grid search entails training and validating the model with assorted hyperparameter combinations within specified ranges, culminating in the selection of the optimal combination via validation set performance. In contrast, cross-validation involves segmenting the training set into multiple folds, training the model on each fold, validating on the remaining folds, and averaging performance metrics to temper evaluation randomness and bolster generalization proficiency. It is prudent to acknowledge that grid search may dictate considerable computational resources and time, mandating judicious hyperparameter range selection and prudent resource allocation to streamline effective hyperparameter tuning [22,23].

LSTM models typically encompass input layers, LSTM layers, and output layers, among other constituents. Model structure can be tailored to data attributes by adjusting parameters such as the number of LSTM neurons and activation functions. During the model training process, setting appropriate hyperparameters—such as learning rate and batch size—assumes significance. Learning rate governs the magnitude of weight updates per iteration, with extremes preventing convergence or inducing local optima. Batch size dictates the number of samples per parameter update, with excessively large batches causing aggressive updates, while overly small batches yield unstable adjustments. Pragmatic
experimentation and optimization are indispensable to ascertain suitable hyperparameter values, fostering superior model performance.

In this work, an LSTM model was crafted within the TensorFlow-GPU 2.9 framework. This model comprises three layers: an input layer, an LSTM layer with 50 neurons; a subsequent LSTM layer with 80 neurons; and the ultimate output layer, featuring a single fully connected neuron for prediction output. Sample data from the past 30 time periods are used to predict data for the next 1 time period. A dropout layer, characterized by a dropout rate of 0.2, intervenes between the second and third layers, systematically discarding a fraction of neuron outputs during model training, thus tempering overfitting risks.

2.4.2. Model Evaluation

The assessment of a model’s predictive performance holds paramount importance in affirming its efficacy. Appropriate evaluation metrics must be used to quantitatively gauge the model’s predicted outcomes. In this study, the mean square error (MSE) and coefficient of determination ($R^2$) emerge as primary indices to scrutinize the predictive prowess of the model [39]. Furthermore, the average absolute error (MAE) and root mean square error (RMSE) are also invoked, furnishing a holistic comprehension of the model’s predictive capacity pertaining to ammonia nitrogen concentration [40–42].

These four evaluation methods are briefly introduced as follows:

1. **Mean square error (MSE):** MSE encapsulates the average of squared differences between predicted values and actual values. It provides a measure of prediction accuracy, with lower MSE values denoting enhanced precision in the model’s predictions.

2. **Coefficient of determination ($R^2$):** $R^2$ quantifies the proportion of the variability in the dependent variable that can be explicated by the model. It ranges from 0 to 1, with higher $R^2$ values indicating stronger model performance in explaining the variance in the data.

3. **Average absolute error (MAE):** MAE computes the average absolute differences between predicted values and actual values. MAE offers insights into the average prediction error magnitude, with lower MAE values reflecting superior prediction accuracy.

4. **Root mean square error (RMSE):** RMSE calculates the square root of the average of squared prediction errors. It provides an estimation of the model’s predictive error spread, with smaller RMSE values signifying improved prediction precision.

By leveraging these evaluation metrics, the model’s performance in forecasting ammonia nitrogen concentration can be rigorously assessed, affording a comprehensive understanding of its predictive capabilities.

2.4.3. Model Optimization

In the realm of model optimization, the consideration of model interpretability assumes significance. Deep learning models are often perceived as “black-box” entities, challenging the explanation of the rationale behind their predictions. To address this challenge, visualization techniques and feature importance analysis can be harnessed to unveil the model’s prediction process. This augments model interpretability, streamlining model application and refinement. It is imperative to recognize that model evaluation and optimization represent iterative processes. Depending on the context, multiple cycles of evaluation and optimization may be warranted, entailing continuous adjustments to model design and parameters until the desired performance benchmarks are met.

In this study, optimization efforts entailed the utilization of grid search and cross-validation methodologies. The model was encapsulated as a regressor via KerasRegressor, thereby enabling its seamless integration with scikit-learn. A GridSearchCV object was instantiated to orchestrate grid search and cross-validation within the designated parameter space. This parameter space encompassed batch size, epochs, and the optimizer. The “cv” parameter dictated the number of folds for cross-validation, set to 2 in this instance, indicating deployment of 2-fold cross-validation [43–45]. After rigorous experimental
comparisons, the following hyperparameters were judiciously selected: a batch size of 32, 50 epochs, and a RMSprop optimizer (root mean square propagation).

RMSprop serves as an optimization algorithm for training neural network models. Operating as an adaptive learning rate technique rooted in the gradient descent algorithm, RMSprop leverages exponentially weighted moving averages of gradients to dynamically adjust the learning rate. In contrast to conventional gradient descent approaches, RMSprop employs the moving average of squared gradients to modulate the learning rate. The central steps of RMSprop entail:

1. Parameter initialization: Weights of the model and exponentially weighted moving average of squared gradients are initialized.
2. Iterative training:
   - Gradients of the model’s loss function concerning the weights are computed.
   - The exponentially weighted moving average of squared gradients is updated.
   - Adjustment value for the learning rate is computed based on the moving average.
   - Weights are updated based on the learning rate adjustment value and gradients.
   - The above steps are reiterated until a termination criterion is satisfied, such as reaching the maximum number of iterations or convergence of the loss function.

RMSprop brings forth several merits, including:

1. Adaptive learning rate: RMSprop dynamically tunes the learning rate in response to gradient changes. Large gradients prompt diminished learning rates, curbing parameter updates, while smaller gradients engender augmented learning rates, hastening parameter updates.
2. Applicability to non-stationary data: RMSprop excels in scenarios with non-stationary gradients, augmenting model training stability and convergence pace.
3. Ameliorating gradient explosion and vanishing: Through the utilization of exponentially weighted moving averages of gradients, RMSprop mitigates the adverse effects of gradient explosion and vanishing, thereby amplifying model training effectiveness.

It remains pivotal to acknowledge that RMSprop mandates manual hyperparameter configuration, including of the initial learning rate and decay coefficient. Additionally, RMSprop may not universally serve as the optimal optimization algorithm, and alternatives such as Adam or Adagrad could outperform RMSprop for specific problems [46–48].

3. Results

3.1. Analysis of Spatiotemporal Variation in NH₃-N Content in River Water Quality

Figure 4 illustrates the fluctuations in NH₃-N concentrations within the Qianshan River. The average NH₃-N concentration follows a discernible diurnal rhythm, culminating in the early morning hours (4:00–08:00) and ebbing during the afternoon (16:00–20:00). This diurnal oscillation can be attributed to the urban lifestyle rhythm. The morning surge in NH₃-N concentration arises from activities such as waking and personal hygiene, which augment the discharge of organic wastewater, subsequently elevating NH₃-N levels. Conversely, afternoon hours, dedicated to work and studies, witness a reduction in organic wastewater discharge, thereby leading to a decline in NH₃-N concentration. Temperature variations between these periods may further contribute. Nighttime features lower water temperatures, which retard microbial metabolic activities, facilitating NH₃-N accumulation. Daytime warmth, in contrast, accelerates microbial metabolism, promoting NH₃-N consumption.

Furthermore, the sway of photosynthesis emerges as a potential influence on NH₃-N fluctuations. Aquatic phytoplankton, through photosynthesis, convert carbon dioxide and water into organic matter and oxygen. This process necessitates NH₃-N and other inorganic nitrogen compounds, thereby ushering a dip in NH₃-N concentration during robust photosynthetic phases in daylight. Subsequently, the absence of photosynthesis during nighttime leads to increased NH₃-N concentration.
Figure 4. Temporal variations of NH$_3$-N within the study area at different time scales.

The daily average NH$_3$-N concentration typically registers an elevation during the middle and upper segments of each month, peaking around the 14th and 15th, while receding during the middle and lower segments, hitting lows around the 18th and 19th. This pattern is intricately intertwined with pollutant emissions and environmental elements. These segments mark peaks for domestic and industrial water usage, leading to wastewater discharge bearing higher NH$_3$-N content and correspondingly elevated NH$_3$-N concentration. Towards the end of the month, as environmental factors and pollutant sources dwindle, NH$_3$-N concentration also gradually diminishes.

Monthly NH$_3$-N concentration averages tend to surge in August and dip in April. This phenomenon likely stems from temperature and climatic alterations. Summer temperatures expedite water chemical reactions, spur bacterial proliferation, and yield additional NH$_3$-N through organic matter decomposition, culminating in heightened NH$_3$-N concentration.
Conversely, spring’s cooler temperatures deter chemical reactions and bacterial growth, translating to decreased NH$_3$-N concentration. Further factors such as increased summer temperatures, reduced rainfall, and slower water flow fostering biological growth and heightened microbial metabolic activity play a role in augmenting NH$_3$-N concentration. Spring’s lower temperatures, amplified rainfall, and swifter water flow, conversely, engender a decline in NH$_3$-N concentration.

The distinct NH$_3$-N concentration trends across varied time spans underscore its cyclic variations in the Qianshan River. The multifaceted factors influencing NH$_3$-N concentration warrant comprehensive consideration for the formulation of effective management strategies against NH$_3$-N pollution. Moreover, these analytical insights provide pivotal reference points, guiding the development, forecasting, and refinement of subsequent deep learning models.

3.2. Evaluation of NH$_3$-N Prediction Performance Based on the LSTM Model

The effectiveness of the developed NH$_3$-N concentration model was rigorously evaluated through the application of key metrics, namely R$^2$, MSE, and MAE, which were all applied to the validation dataset. The outcomes of this evaluation validate the proficiency of the LSTM model within the research domain, as depicted in Figure 5.

![Figure 5](image.png)

Figure 5. Learning curves and prediction results of the LSTM model on the validation dataset, along with the corresponding R$^2$ values: (a) learning curve (MAE); (b) learning curve (MSE); (c) observed and predicted NH$_3$-N concentrations, along with the corresponding R$^2$ value.
The model’s trajectory of convergence and stability was observed within a span of 50 iterations. This achievement was coupled with an impressively low MAE that remained below 0.045, accompanied by a MSE that maintained itself below 0.004. This portrayal in Figure 5 succinctly underscores the model’s efficacy in predicting ammonia nitrogen concentrations. The proximity of these metrics to their respective minima enforces the LSTM model’s competence in forecasting ammonia nitrogen concentration.

Furthermore, the predictive outcomes gleaned from the model, as aptly showcased in Figure 5c, manifest a remarkable alignment with the actual measured values. This agreement is further underscored by the calculated R$^2$ value of 0.89. In totality, the LSTM model deftly captures the nuanced concentration variations of NH$_3$-N coursing through the Qianshan River, thus emerging as a robust and adept predictive model.

3.3. Comparison of NH$_3$-N Prediction Performance Based on Different Feature Sets

In order to identify the key input variables combinations that influence the prediction results of ammonia nitrogen concentrations, the LSTM model was utilized with different combinations of the nine input variables to predict ammonia nitrogen levels on the test dataset. Based on the strength of the correlation between the input variables and the target output, the nine input variables were sorted in descending order of their correlation coefficient values with the target output. The input feature combinations were gradually formed by cumulatively adding the correlation coefficient values, as shown in Figure 3.

In the sphere of evaluation metrics, the R-squared (R$^2$) value emerged as a cardinal yardstick, affording substantive insights into the model’s capacity to explain the target variable. Spanning the continuum from 0 to 1, an R$^2$ value approaching unity connoted heightened explanatory efficacy of the model relative to the target variable. Our meticulous scrutiny of R$^2$ values unveiled the preeminence of feature combination 6, a composition encompassing six variables, which secured the acme R$^2$ value of 0.82. This pronounced R$^2$ value underscored the compelling explanatory prowess wielded by feature combination 6 over the target variable.

Furthermore, our scrutiny extended to mean squared error (MSE) and root mean squared error (RMSE), metrics poised to gauge the dissonance between the model’s prognostications and the empirical observations. Remarkably, feature combination 6 evidenced commendable proficiency, yielding nominal error values of 0.0047 and 0.0655 for MSE and RMSE, respectively, thereby accentuating the model’s prowess in delivering refined predictive accuracy.

Concomitantly, the focus converged on mean absolute error (MAE), a barometer of the average absolute divergence between the model’s prognoses and the actual observations. In this purview, feature combination 6 preserved its ascendancy, manifesting a modest absolute error value of 0.0460, an indication of its robust capacity to attenuate prediction bias.

A comprehensive synthesis of Figure 6 and Table 3 unveils compelling revelations. Feature combination 1, characterized by a single indicator, boasted an elevated R$^2$ value of 0.79, alongside mitigated MSE, RMSE, and MAE values. This configuration accentuated the salience of a single feature’s explanatory potential with regard to the target variable, indicating a heightened predictive accuracy. In contrast, feature combinations 2, 3, 4, and 7 followed a trajectory marked by diminished R$^2$ values and accentuated MSE, RMSE, and MAE values—reflective of dwindling explanatory efficacy and curtailed predictive precision. Feature combinations 5, 8, and 9 presented consistent performance, exhibiting amplified R$^2$ values juxtaposed against marginally inflated MSE, RMSE, and MAE values with regard to feature combination 6. This nuanced differentiation intimates a marginal reduction in predictive accuracy for these configurations.

A holistic assimilation of the aforesaid analysis unequivocally elevates feature combination 6—a composite of six variables—as an exemplar of superlative performance, as evidenced across a spectrum of evaluation metrics. Supported by an improved R$^2$ value, reduced MSE, RMSE, and MAE values, as well as enhanced predictive precision, feature
combination 6 emerges as a compelling set of input variables, deserving of thorough investigation in future research initiatives and practical implementations.

Figure 6. Comparison of observed and predicted NH$_3$-N concentrations on different feature sets on the test dataset, along with the corresponding R$^2$ values: (a) prediction results of combination 1; (b) prediction results of combination 2; (c) prediction results of combination 3; (d) prediction results of combination 4; (e) prediction results of combination 5; (f) prediction results of combination 6; (g) prediction results of combination 7; (h) prediction results of combination 8; (i) prediction results of combination 9.
Table 3. Summary statistics of feature sets and their corresponding evaluation metric values.

<table>
<thead>
<tr>
<th>Indicator Combination</th>
<th>$R^2$</th>
<th>MSE</th>
<th>RMSE</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combination 1</td>
<td>0.79</td>
<td>0.0045</td>
<td>0.0674</td>
<td>0.0453</td>
</tr>
<tr>
<td>Combination 2</td>
<td>0.74</td>
<td>0.0053</td>
<td>0.0730</td>
<td>0.0503</td>
</tr>
<tr>
<td>Combination 3</td>
<td>0.73</td>
<td>0.0048</td>
<td>0.0695</td>
<td>0.0518</td>
</tr>
<tr>
<td>Combination 4</td>
<td>0.75</td>
<td>0.0052</td>
<td>0.0722</td>
<td>0.0543</td>
</tr>
<tr>
<td>Combination 5</td>
<td>0.75</td>
<td>0.0067</td>
<td>0.0816</td>
<td>0.0629</td>
</tr>
<tr>
<td>Combination 6</td>
<td>0.82</td>
<td>0.0047</td>
<td>0.0655</td>
<td>0.0460</td>
</tr>
<tr>
<td>Combination 7</td>
<td>0.75</td>
<td>0.0066</td>
<td>0.0814</td>
<td>0.0630</td>
</tr>
<tr>
<td>Combination 8</td>
<td>0.72</td>
<td>0.0081</td>
<td>0.0900</td>
<td>0.0695</td>
</tr>
<tr>
<td>Combination 9</td>
<td>0.78</td>
<td>0.0057</td>
<td>0.0752</td>
<td>0.0576</td>
</tr>
</tbody>
</table>

4. Discussion

4.1. Feasibility of Using the LSTM Model with Easily Measurable Water Quality Data to Predict Ammonia Nitrogen Concentrations in Water Quality

The LSTM model developed in this study has effectively established a nonlinear mapping relationship between readily measurable water quality parameters ($NH_3$-N, Temp, pH, DO, K$_{MnO_4}$, TP, TN, Cond, and Turb) and the target variable ($NH_3$-N). This achievement helps to accurately predict the concentration of ammonia nitrogen in river systems. The model’s predicted $NH_3$-N concentrations closely align with observed values acquired from real-time data collected at river water monitoring sites, attaining an average $R^2$ value of 0.82. This reflects the strong ability of the model to predict the peak concentration of $NH_3$-N, which can provide reliable early warnings to mitigate the impact of elevated $NH_3$-N levels on water quality. This ability is of great significance for intelligent monitoring and management of aquatic environments.

It is worth noting that, unlike the accuracy of predicting concentration peaks, the model exhibits a slight decrease in its effectiveness in predicting $NH_3$-N concentration valleys within specific time intervals, as shown in Figure 6c–i. The model did not fully learn the complexity of data attributes during the training phase $NH_3$-N concentration trough, making it impossible to accurately predict the trough value. Moreover, the limited $NH_3$-N concentrations within water quality samples during valley periods, coupled with potential measurement deviations in Internet of Things (IoT) real-time monitoring devices, may engender diminished accuracy in raw data. These exceptional circumstances inevitably contribute to the attenuation of training data accuracy, consequently influencing the model’s predictive performance. Hence, the acquisition of high-fidelity training data assumes critical importance to reinforce prognostic precision. Furthermore, predicated upon the findings presented in Figure 6, the incorporation of supplementary input variables that wield substantive influence over the output variable could potentially ameliorate model prediction accuracy. Thus, delving into additional potential indicators that affect $NH_3$-N concentrations within river water quality can serve to augment the model’s efficacy, presenting a meaningful avenue for enhancing predictive capabilities.

4.2. Potential for Reducing Model Prediction Costs

In contrast to conventional mechanistic models, the data-driven prediction methodologies surmount the temporal limitations associated with sample procurement, analysis, and detection, concurrently reducing the demand for substantial human, financial, and material resources. However, when scrutinizing indicators that are measurable within brief time-frames of minutes to a day, the adoption of sensors having high temporal resolution might entail elevated costs, notably in terms of instrument probe maintenance. Thus, it is imperative to identify pivotal variables for model training that exhibit a minimal compromise on prediction performance. This understanding may significantly improve the operational efficiency of the model, thereby reducing computational energy consumption and prediction expenses. Empirical findings indicate that using a separate input–output paradigm in the prediction model to achieve accurate $NH_3$-N prediction is not sufficient. Additionally, the
iterative approach of progressively augmenting input variables to discern the optimal input combination yielding superior NH$_3$-N prediction accuracy entails notable temporal and operational investments. Conversely, the methodological application of Pearson correlation coefficient analysis effectively identifies a subset of input variables characterized by robust interactions that materially contribute to the model’s output. Notably, the current study demonstrates the relevance of NH$_3$-N, pH, DO, KMnO$_4$, TP, and TN (as delineated in Figure 3). Therefore, the composition of input variables is amenable to adjustment contingent upon the ordering of their correlations, thereby engendering the identification of an optimal input indicator combination predicated upon prediction performance.

5. Conclusions

In this investigation, a data-driven Long Short-Term Memory (LSTM) model was designed to predict NH$_3$-N concentrations in river water networks. This model shows good accuracy in predicting NH$_3$-N concentration. Primarily, an exploratory examination was undertaken to assess the aptitude of deep learning methodologies in NH$_3$-N prediction. The outcomes manifestly underscored the data-driven NH$_3$-N prediction model’s robust generalization potential, led by an impressive $R^2$ value of 0.82 for the optimal input indicator amalgamation. Furthermore, the model’s performance was amenable to enhancement through judiciously modulating the number of layers and neurons within the LSTM framework. Equally noteworthy, the employment of Pearson correlation coefficient analysis expeditiously illuminated and quantified the multi-faceted contributions of diverse input variables to the model’s predictive outcomes. This analytical framework significantly enriched our comprehension of deep learning results and facilitated model optimization. Overall, the proposed LSTM-based NH$_3$-N prediction model effectively overcomes the limitations of traditional monitoring methods in terms of time and economic costs and enables fast modeling at low costs. This provides a feasible solution for early warning of high NH$_3$-N concentrations in river water, enabling water environmental management departments to develop inspection plans and reduce incidents of water quality hazards caused by excessive NH$_3$-N concentrations.

However, the proposed model has some limitations. Rooted in the underpinnings of the deep learning algorithm, modeling efficacy hinges upon the interplay between input and output variables. This study predominantly accentuated the correlation existing between input and output variables, thereby inadvertently disregarding the latent interplay amongst the input variables themselves. This analytical disposition could potentially lead to the inadvertent omission of pivotal variables, given the plausible existence of inherent correlations amongst the input variables. It is plausible that certain features might furnish supplementary insights that underpin enhanced predictive capabilities. The inadvertent oversight of internal feature correlations could yield the exclusion of such salient features, impinging upon the model’s precision and performance. Additionally, the introduction of redundant features—features with a high degree of correlation—might entail unwarranted complexities, hampering model training and generalization proficiencies. In scenarios wherein inter-feature correlations exist, the model may inadvertently assign disproportionate weights to features exhibiting elevated correlations, inadvertently sidelining features characterized by lower correlations. This asymmetry in feature weighting can engender biased feature attributions and potentially compromise the model’s adeptness in harnessing the complete spectrum of available information. Neglecting the intrinsic feature correlations further augments the model’s tendency to disproportionately depend on specific features during the training phase, thereby amplifying the risk of overfitting. Consequently, in the realm of feature engineering, a judicious consideration of both the correlation between input–output variables and the internal inter-feature correlations is indispensable, potentially culminating in more exhaustive and precise predictive models.

In summary, the future research should focus on improving model performance, expanding application domains, streamlining workflows, and further enhancing model interpretability to better support various aspects of water quality environmental man-
management and governance. To better understand the model’s performance variations at different times, we plan to incorporate seasonality and other temporal patterns as input features. This step will enable us to more accurately capture the seasonal variations in water quality, providing precise data support for water quality management. Furthermore, we will actively investigate various data processing and feature selection methods, such as principal component analysis and causal analysis, to gain a deeper understanding of the reasons behind performance differences. By continually optimizing the model, enhancing its generalization capabilities and robustness, we will ensure that the model performs excellently under diverse conditions [49]. Simultaneously, we will compare the performance of different deep learning models, streamline model algorithms, and improve model interpretability. By quantifying model costs, we will maintain efficient workflows while enhancing performance. This approach will better serve the needs of water quality prediction and water quality environmental management.

**Author Contributions:** Conceptualization, X.W., Y.L. (Yanchun Liang) and A.T.; methodology, M.Q.; software, X.W. and Y.L. (Ying Li); validation, A.T., X.W. and M.Q.; formal analysis, X.W. and A.T.; investigation, X.W.; resources, M.Q. and Q.Q.; data curation, X.W. and Y.L. (Ying Li); writing—original draft preparation, X.W.; writing—review and editing, Q.Q., A.T. and Y.L. (Yanchun Liang); visualization, X.W.; supervision, Y.L. (Yanchun Liang); project administration, Y.L. (Ying Li). All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded in part by the NSFC (62372494, 61972174), the Guangdong Universities’ Innovation Team (2021KCXTD015), the Key Disciplines Projects (2021ZDJS138), and the Guangdong Provincial Junior Innovative Talents Project for Ordinary Universities (2022KQNCX146).

**Data Availability Statement:** Data shall be provided by the first and corresponding authors upon special request.

**Acknowledgments:** Thanks to Qingyue Data (data.epmap.org) for support of environmental data.

**Conflicts of Interest:** The authors declare no conflict of interest.

**References**


22. Jiang, Y.; Li, C.; Song, H.; Wang, W. Deep Learning Model Based on Urban Multi-Source Data for Predicting Heavy Metals (Cu, Zn, Ni, Cr) in Industrial Sewer Networks. J. Hazard. Mater. 2022, 432, 128732. [CrossRef]


34. Zavareh, M.; Maggioni, V. Application of Rough Set Theory to Water Quality Analysis: A Case Study. Data 2018, 3, 50. [CrossRef]


42. Hadjiisolomou, E.; Stefanidis, K.; Herodotou, H.; Michaelides, M.; Papatheodorou, G.; Papastergiadou, E. Modelling Freshwater Eutrophication with Limited Limnological Data Using Artificial Neural Networks. Water 2021, 13, 1590. [CrossRef]


44. Ahmad, G.N.; Fatima, H.; Ullah, S.; Saidi, A.S. Efficient Medical Diagnosis of Human Heart Diseases Using Machine Learning Techniques with and without GridSearchCV. IEEE Access 2022, 10, 80151–80173. [CrossRef]


47. Kumar, A.; Sarkar, S.; Pradhan, C. Malaria Disease Detection Using Cnn Technique with Sgd, Rmsprop and Adam Optimizers. In Deep Learning Techniques for Biomedical and Health Informatics; Springer: Cham, Switzerland, 2020; pp. 211–230. [CrossRef]


49. Zavareh, M.; Maggioni, V.; Sokolov, V. Investigating Water Quality Data Using Principal Component Analysis and Granger Causality. Water 2021, 13, 343. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.