Effective Short-Term Continuous Data Prediction Using LSTM-NHITS

Xiaohong Peng[1], Tianrong Zhong[1], Renyou Yang[2], Zhao Li[1]*

[1] School of Mathematics and Computer Science, Guangdong Ocean University, China
[2] Southern Marine Science and Engineering Guangdong Laboratory (Zhanjiang), China
lgdpxh@126.com, ztrgdufe@163.com, youngrenyou@zjblab.com, zhaoli@gdou.edu.cn

Abstract

Accurate short-term continuous data prediction is crucial for timely water quality assessment and pollution prevention. However, the nonlinear and temporally dependent nature of water quality data presents significant challenges for traditional forecasting models. To address these challenges, we propose an effective short-term continuous prediction model, LSTM-NHITS, which combines Long Short-Term Memory (LSTM) networks with Neural Hierarchical Interpolation for Time Series (NHITS). This model effectively captures multi-scale features and complex temporal dependencies, improving prediction accuracy. Experimental results from datasets collected at multiple monitoring stations in Zhanjiang City show that LSTM-NHITS outperforms traditional models across different short-term forecasting horizons (4-hour, 12-hour, and 1-day). By accurately modeling both long- and short-term dependencies, this approach ensures precise continuous water quality prediction, demonstrating its potential for real-time environmental monitoring and management.

Keywords: Surface water quality prediction, Multi-level attention mechanism, LSTM-NHITS, Short-term prediction, Zhanjiang City

1 Introduction

As industrialization and urbanization accelerate, monitoring surface water quality has become increasingly important due to rising water pollution levels. Effective monitoring helps authorities ensure that water meets national standards, thereby protecting public health. Changes in water quality affect ecosystem balance, human life, and species survival, making accurate prediction and effective management strategies essential.

Government agencies and research institutions have been leveraging modern technologies to improve the accuracy and real-time performance of water quality predictions. Advances in monitoring equipment, along with IoT and big data, have greatly enhanced data collection and processing capabilities. Deep learning models, in particular, have proven effective for rapid water quality analysis and prediction [1].

However, several challenges remain in practical applications:

1. Data sparsity: Limited monitoring points and insufficient spatial coverage result in missing data, reducing the reliability of predictions.

2. Complexity of monitoring regions: Geographical and environmental heterogeneity complicates prediction efforts.

3. Limitations of existing models: Current methods struggle to capture the temporal and nonlinear characteristics of water quality dynamics, impacting prediction accuracy and stability.

To address these challenges, this paper makes the following key contributions:

1. Mitigating data sparsity: A novel data fusion strategy integrates multi-source data to address missing data and improve model reliability.

2. Accounting for regional complexity: A hybrid deep learning and traditional forecasting approach enhances prediction accuracy and stability across different regions.

3. Enhancing model capabilities to capture dynamic features: The optimized LSTM-NHITS model combines temporal and nonlinear dynamics with a multi-level data structure, improving both prediction accuracy and stability.

These innovations provide a solid foundation for advancing water quality monitoring and management strategies.

2 Related Works

Water quality prediction has been extensively studied to improve accuracy and real-time performance while addressing challenges like data sparsity, complex monitoring regions, and the limitations of existing models in capturing dynamic water quality changes.

1. Data Acquisition Sparsity

Early methods relied on traditional statistical approaches. Zuo et al. [2] proposed a VMD-TCN-ARIMA framework optimized with the weighted whale optimization algorithm (WSWOA) to handle time-related fluctuations, though its accuracy suffered from data sparsity. To address this, Nsabimana et al. [3] applied a gray Markov model to improve groundwater hardness prediction, enhancing data reliability and addressing sparsity.

2. Complexity of Monitoring Regions

Traditional statistical models struggled with the

*Corresponding Author: Renyou Yang; Email: youngrenyou@zjblab.com
DOI: https://doi.org/10.70003/160792642026052703011

complexity of monitoring regions. Zhang et al. [4] combined LS-SVM with PSO to handle linear data, but this approach had limitations for complex water quality changes. Deep learning models like LSTM have been introduced to better handle non-linear data [5-7]. For example, Sankarasubbu et al. [8] enhanced LSTM for real-time monitoring, and Gauch et al. [9] developed an MTS-LSTM model for multi-scale rainfall-runoff prediction, improving monitoring in complex environments.

3. Capturing Dynamic Changes in Water Quality

Existing models often fail to capture dynamic, non-linear water quality variations. Zhang et al. [10] proposed the LSTMAE-XGBOOST model, combining autoencoders, LSTM, and XGBOOST for improved accuracy. Other models, like the hybrid Bayesian-LSTM-GRU introduced by Wang et al. [11] and attention-driven LSTM (A-LSTM) and A-GRU models by Gandh et al. [12], focus on dynamic variations, improving efficiency in aquaculture applications.

Despite advancements, challenges remain, including dataset enrichment, model optimization, and computational efficiency. Generalization is also an issue, particularly for models tailored to specific regions. To address these issues, this study proposes a hybrid LSTM-NHITS model to improve accuracy and effectiveness in water quality prediction, overcoming key limitations in existing research.

3 Methods and Model Improvements

3.1 Identification of Existing Issues

Water quality monitoring and prediction face several critical challenges, which limit the effectiveness and reliability of existing methods. These issues can be summarized as follows:

1. Insufficient Capability to Handle Complex Nonlinear Relationships: Traditional models struggle with nonlinear relationships in water quality data, leading to large prediction errors, expressed as:

$$Error = | y_{ture} - y_{pred} | \tag{1}$$

y_{ture} and y_{pred} represent the true and predicted values, respectively.

2. Data Sparsity Issues: Insufficient real-time monitoring results in limited training samples, particularly in regions with inadequate monitoring infrastructure. The availability of samples can be represented as:

$$N = | D \geq T \tag{2}$$

N represents the available sample size, $|D$ is the total number of samples in the dataset, and T is the minimum required number of training samples. When $N < T$, the training process becomes challenging, and the prediction performance of the model is adversely affected.

3. Poor Model Generalization Ability: Existing models often fail to generalize well, leading to poor performance across different water bodies or environmental conditions. This can be quantified as:

$$\mathcal{L}_{\text{train}} \approx \mathcal{L}_{\text{test}} \tag{3}$$

$\mathcal{L}_{\text{train}}$ represents the training loss, and $\mathcal{L}_{\text{test}}$ denotes the testing loss. Poor generalization leads to overfitting, causing performance drops on unseen data.

To address these challenges, this study introduces a hybrid method combining Long Short-Term Memory (LSTM) networks with Neural Hierarchical Interpretable Time Series (NHITS) models. The LSTM component captures long-term temporal dependencies, while NHITS improves accuracy and robustness through hierarchical feature extraction and multi-scale handling of nonlinear dynamics.

This LSTM-NHITS hybrid method overcomes the limitations of traditional models, enhancing prediction accuracy, generalization, and robustness in sparse, heterogeneous datasets and diverse environmental conditions.

3.2 Overall Framework

This study proposes a hybrid water quality prediction model that integrates the Long Short-Term Memory (LSTM) network [13] and the Neural Hierarchical Interpretable Time Series (NHITS) network [14]. By leveraging multi-level feature extraction and processing, the framework effectively addresses key challenges in water quality monitoring, including data complexity, data sparsity, and insufficient model generalization.

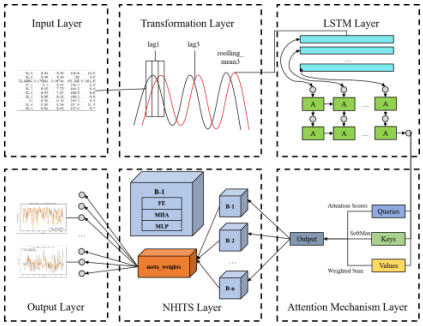

Figure 1. Framework overview

The structure of the proposed LSTM + NHITS network is illustrated in Figure 1. The model is composed of six main layers: the input layer, transformation layer, LSTM layer, attention mechanism layer, NHITS layer, and output layer. The functions and interactions of each layer are described as follows:

3.2.1 Input Layer

The input layer processes multi-scale time series data, encompassing water quality measurements over different time spans. In this study, the input includes time series with 12, 36, and 72 steps, where each time step contains *features_dim* features. The multi-dimensional time series can be mathematically expressed as:

$$X_t = \{X_{12}, X_{36}, X_{72}\} \in \mathbb{R}^{T \times features\_dim} \tag{4}$$

T represents the number of time steps, and *features_dim* denotes the number of features at each time step. The

use of multi-scale time series enables the model to capture both long-term trends and short-term fluctuations in water quality changes, providing comprehensive temporal information for prediction.

3.2.2 Transformation Layer

The transformation layer preprocesses the data from the input layer to ensure consistency across different time scales. Normalization and other feature transformation techniques are applied to reduce noise and standardize the input data. This layer also distributes the preprocessed data to different processing paths, optimizing the input for the subsequent LSTM and NHITS layers. By harmonizing data formats and directing data appropriately, the transformation layer ensures that the multi-scale information is efficiently utilized for feature extraction and prediction.

3.2.3 LSTM Layer

The LSTM layer serves as the core of the proposed model, capturing long-term dependencies in water quality time series. While a unidirectional LSTM is employed in this study, a bidirectional LSTM (Bi-LSTM) could be explored in future work to enhance sequential feature learning. By processing sequences in both forward and backward directions, Bi-LSTM can capture richer temporal dependencies. However, given that NHITS already performs multi-resolution decomposition, additional forward-backward dependencies might introduce computational redundancy. Future studies could evaluate whether Bi-LSTM improves short-term prediction accuracy, especially in scenarios with rapid water quality fluctuations.

LSTM networks are particularly effective at modeling temporal patterns influenced by climatic variations and human activities. Each input sequence is processed sequentially through LSTM units, where gating mechanisms (input gate, forget gate, and output gate) regulate information flow. The hidden states are updated at each time step, capturing critical dependencies in water quality trends and providing essential support for subsequent layers. The LSTM transformation is represented as:

$$h_t = f\left(W_h h_{t-1} + W_x x_t + b\right) \quad (5)$$

h_t denotes the hidden state at time step, X_t is the input, and W_h, W_x represent the model parameters. This mechanism ensures effective memory retention, strengthening the predictive performance of the overall framework.

3.2.4 Attention Mechanism Layer

The attention mechanism layer weights the temporal features extracted by the LSTM layer, assigning dynamic importance to different time steps. This adaptive weighting highlights critical moments in the time series, enhancing the model's responsiveness to significant changes. The attention mechanism computes a weight matrix α_t for each time step and performs a weighted average on the LSTM outputs, as represented by:

$$\hat{h}_t = \alpha_t \cdot h_t \quad (6)$$

α_t is the learned weight for time step t, and $\hat{h}_t$ is the LSTM output at that time step. By adaptively focusing on relevant time periods, the attention mechanism improves the model's ability to handle dynamic changes in water quality, ensuring more accurate predictions.

3.2.5 LSTM-NHITS Interaction Mechanism

To integrate LSTM with NHITS, an adaptive feature transition mechanism is introduced, converting sequential modeling into hierarchical decomposition. The final hidden states from different time spans (e.g., 12, 36, and 72 steps) are concatenated to form a multi-resolution feature representation:

$$H_{agg} = [h_{T-12}, h_{T-36}, h_{T-72}] \quad (7)$$

This aggregated feature enables the model to capture both short-term fluctuations and long-term trends in water quality. A self-attention mechanism is applied to dynamically assign importance weights to time steps, ensuring critical temporal dependencies are emphasized. This attention-based feature selection process is formulated as:

$$H_{attn} = \sum_t \alpha_t \cdot h_t, \alpha_t = softmax\left(W_\alpha h_t\right) \quad (8)$$

After attention refinement, the selected features are projected into NHITS's hierarchical framework for time series decomposition:

$$X_{NHITS} = W_p H_{attn} \quad (9)$$

This projection allows NHITS to separate high-frequency and low-frequency components, improving the model's ability to handle both transient anomalies and long-term patterns. By leveraging the complementary strengths of LSTM and NHITS, this interaction mechanism enhances predictive performance, ensuring accurate and robust water quality forecasts across temporal scales.

3.2.6 NHITS Layer

The NHITS layer refines features extracted by the LSTM and attention layers through hierarchical feature extraction and weighting mechanisms. It consists of multiple NHITS blocks, each with three components: feature extraction, multi-head attention, and a multilayer perceptron (MLP).

Figure 2 illustrates the architecture of a single NHITS block, which processes input features in the following steps:

1. Feature Extraction (FE): The feature extraction layer applies linear and nonlinear transformations to capture underlying patterns in the time series. This is expressed as:

$$F_k = \sigma\left(W_F X_k + b_F\right) \quad (10)$$

F_k represents the feature vector at the k-th time step after the transformation, W_F is the weight matrix for the feature extraction transformation, X_k is the input feature vector at the k-th time step, b_F is the bias term, and σ is a nonlinear activation function, such as the sigmoid or ReLU function, that applies a nonlinear transformation to the output.

2. Multi-Head Attention Mechanism (MHA): This mechanism computes query, key, and value matrices from the features, capturing multi-scale temporal dependencies. The attention operation is:

$$Attention(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (11)$$

Q, K, and V are the query, key, and value matrices, and d_k is the key vector dimension. This operation assigns dynamic attention weights to the input features.

3. Multilayer Perceptron (MLP): The MLP layer applies nonlinear transformations to the attention output to generate intermediate outputs. This is defined as:

$$O_k = \sigma(W_O A_k + b_O) \quad (12)$$

O_k represents the output of the MLP at the k-th time step, W_O is the weight matrix applied to the attention output A, and b_O is the bias term.

The output of each NHITS block can be expressed as:

$$y_{NHITS}^{(i)} = MLP\left(MHA\left(FE\left(\hat{h}_t, \theta_{NHITS}^{(i)}\right)\right)\right), i = 1,...,N \quad (13)$$

$\hat{h}_t$ is the input from the attention layer, θ_{NHITS} represents the parameters of the i-th NHITS block, and y_{NHITS} is the block's output. By stacking multiple NHITS blocks, the model achieves hierarchical feature learning, enhancing its robustness to sparse data and improving its ability to capture multi-scale time features.

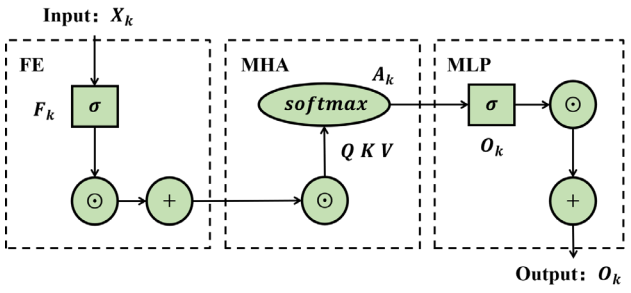

Figure 2. Architecture of a single NHITS block

3.2.7 Output Layer

The output layer aggregates the results from all NHITS blocks to generate final multi-step predictions for different time horizons (e.g., 4 hours, 12 hours, and 1 day). A weighted average strategy is applied to integrate the outputs of NHITS blocks, as represented by:

$$y_{final} = \sum_{i=1}^{N} meta\_weights_i \cdot y_{NHITS}^{(i)} \quad (14)$$

$meta\_weights_i$ denotes the weight of the i-th NHITS block, and is its output. By leveraging the predictive capabilities of each NHITS block, the output layer produces accurate and robust forecasts. This multi-step prediction mechanism ensures that the model effectively handles both short-term and long-term forecasting tasks, providing reliable results for water quality management.

3.3 Data Selection and Preprocessing

3.3.1 Data Selection

Zhanjiang City, located in southwestern Guangdong Province, China, borders the South China Sea and covers approximately 13,225 square kilometers. As a key coastal port city, it is divided into four urban districts, three county-level cities, and two counties. Known for its tropical agriculture and port economy, Zhanjiang's unique location and abundant resources make it vital for agriculture, fisheries, and industry.

The city has a tropical-subtropical climate with an average annual temperature of 23°C and rainfall ranging from 1,400 to 1,900 mm, creating ideal conditions for agricultural and aquatic development. Surface water resources, including rivers and lakes on the Leizhou Peninsula, are crucial for irrigation, fisheries, industry, and domestic use, while also supporting ecological sustainability.

However, rapid urbanization and industrialization have led to challenges in surface water quality, including issues like water eutrophication and ecological imbalance due to agricultural runoff, industrial discharge, and sewage [14]. As a result, maintaining water quality has become a key environmental focus.

This study selects four representative monitoring sites in Zhanjiang:

1. Shan Jiao (SJ) (21°54′43.65″N, 110°17′32.838″E),
2. Chikan Water Plant Intake (CKSC) (21°17′52.764″N, 110°18′37.292″E),
3. Pa Li (PL) (23°01′54.811″N, 115°41′31.003″E),
4. Ying Zai (YZ) (21°53′58.146″N, 111°35′56.432″E).

Water quality data for the study is sourced from the National Surface Water Quality Automatic Monitoring System, covering June 17, 2021, to December 23, 2023, with samples taken every 4 hours. The dataset includes nine key water quality parameters: water temperature, pH, dissolved oxygen, turbidity, conductivity, permanganate index, ammonia nitrogen, total phosphorus, and total nitrogen. These parameters provide comprehensive insights into water quality trends, facilitating detailed analysis and prediction.

3.3.2 Data Preprocessing

Water quality monitoring data often contain missing values due to equipment issues or environmental disturbances, leading to inconsistencies in time series

datasets [15]. Addressing these issues is essential for data accuracy and continuity.

1. Missing Data Handling

Scholars have extensively studied interpolation methods to address missing values in time series data [16-17]. Cubic spline interpolation is employed to handle missing values, leveraging temporal dependencies and correlations among water quality indicators. This approach ensures accurate filling by utilizing long-term trends and interdependencies. The cubic spline interpolation function is:

$$y(t)=\sum_{i=0}^{n} a_i \cdot B_i(t) \tag{15}$$

$y(t)$ is the predicted value at the interpolation point, $B_i(t)$ represents the cubic spline basis functions, and a_i are the coefficients derived from known data points. For boundary data, forward and backward filling techniques are used.

2. Data Standardization

Min-max normalization is applied to standardize the water quality parameters, rescaling all features to a uniform range [18]. The normalization process is expressed as:

$$x'=(x-x_{min})/(x_{max}-x_{min}) \tag{16}$$

x is the original value, x_{min} and x_{max} are the minimum and maximum values in the dataset.

3. Data Cleaning and Interpolation Implementation

Python 3.8 is used to clean non-numeric characters (e.g., '--', '*') by replacing them with NaN values. This cleaning process is expressed as:

$$D'=\begin{cases} D, & if\ D\ is\ numeric \\ NaN, if\ D\ is\ non-numeric \end{cases} \tag{17}$$

D represents the raw data, and D' denotes the cleaned data.

A complete time series, from June 17, 2021, at 12:00 to December 23, 2023, at 16:00, with 4-hour intervals, is generated for missing time points. The total number of time points is:

$$T=\{t_0,t_1,t_2,\ldots,t_n\}, t_i=t_0+4i(i=0,1,2,\ldots,n) \tag{18}$$

t_0 is the starting time, t_n is the ending time, and represents the total number of time points. For time points not included in the original dataset, missing data records are created while retaining basic site information, such as location, monitoring site name, and site conditions.

4. Output and File Saving

After interpolation and normalization, the processed dataset is saved as a CSV file for further model training and analysis:

$$Output = Save(D', format = CSV) \tag{19}$$

Output represents the fully processed dataset. This preprocessing pipeline ensures the dataset is clean, complete, and ready for predictive modeling.

3.3.3 Feature Selection

Surface water quality is a key indicator of ecological health. Unlike closed water bodies, surface water systems are dynamic and self-regulating, though human activities like industrial discharge and agricultural fertilization can significantly impact water quality. These systems often exhibit resilience, but the interplay of natural and anthropogenic factors makes water quality fragile and prone to fluctuations [19].

To assess and monitor surface water quality, this study identifies five key parameters:

1. Water Temperature (T): Affects aquatic life and chemical reactions.

2. pH Value (pH): Reflects water acidity or alkalinity, crucial for ecological balance.

3. Dissolved Oxygen (DO, mg/L): Indicates oxygen availability, vital for aquatic organisms.

4. Turbidity (Turbidity): Represents water clarity and suspended particles, often linked to pollution.

5. Conductivity (EC, μS/cm): Reflects ionic content, associated with dissolved salts and pollutants.

$$WQ = f(T, pH, DO, NTU, EC) \tag{20}$$

WQ denotes the water quality assessment indicator, and *f* represents the functional relationship between the selected parameters and overall water quality. These parameters serve as critical inputs for monitoring and predictive models, enabling effective assessment and management of surface water health.

3.3.4 Model Training

The model training process is illustrated in Figure 3.

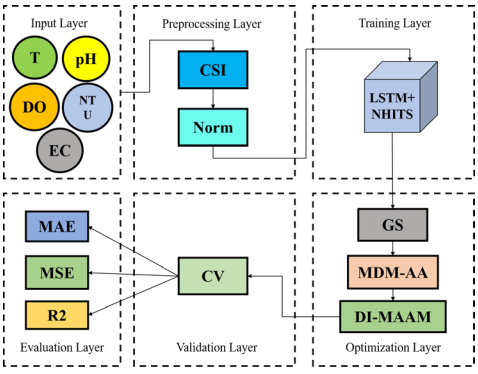

Figure 3. Training flowchart

The training process, shown in Figure 3, employs a hybrid approach combining LSTM networks with NHITS models for effective water quality forecasting. The dataset is split into 80% training and 20% testing sets, with training conducted on an NVIDIA GeForce RTX 3060 GPU. Two strategies were used:

1. Joint training, where LSTM and NHITS update parameters synchronously, optimizing a shared loss function in each iteration.

2. Alternate training, where LSTM and NHITS update parameters separately in alternating steps, allowing each model to adapt to different temporal structures independently.

To enhance efficiency, the training time per epoch is recorded under various batch sizes and learning rates. Experimental results indicate that joint training leads to faster convergence, while alternate training improves adaptability across multiple time scales.

During model selection, grid search (GS) is utilized for hyperparameter tuning, optimizing parameters such as learning rate, batch size, and network depth. The optimization process follows:

$$arg\min_{\theta} \mathcal{L}(\theta) = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \quad (21)$$

θ represents the hyperparameter combination, $\mathcal{L}$ is the loss function (mean squared error), y_i and $\hat{y}_i$ are the actual and predicted values, respectively, and N is the number of samples. Incorporating cross-validation (CV) in the validation layer significantly enhances the stability and accuracy of the model.

For model optimization, parameter synchronization between LSTM and NHITS is a key consideration. In joint training, both models update parameters simultaneously, ensuring consistent gradient propagation. In alternate training, LSTM and NHITS are updated sequentially, which may introduce gradient conflicts where the optimization directions diverge. To mitigate this issue, the following techniques are implemented:

1. Gradient Clipping, to limit excessive updates and improve stability.

2. Adaptive Gradient Blending, to balance the influence of LSTM and NHITS gradients dynamically.

3. Lookahead Optimization, to refine gradient updates and stabilize convergence.

To enhance adaptability, Multiscale Dynamic Modeling and Adaptive Analysis (MDM-AA) adjusts time scales, and Deep Integration of Multilevel Adaptive Attention Mechanism (DI-MAAM) improves feature selection and prediction accuracy. Each time scale is trained for 50 iterations, with the best model weights saved for evaluation.

Performance is evaluated using MAE, MSE, and R^2 metrics, showing that the hybrid model outperforms others in prediction accuracy and robustness, making it ideal for water quality forecasting.

3.3.5 Programming Environment and Training Parameters

For the model training process in this study, Python 3.8 was utilized as the programming environment. The following libraries and tools were employed to facilitate data processing, model construction, and training:

Pandas: This library was used for data processing and management, particularly effective in handling missing data and converting data formats.

NumPy: Utilized for efficient numerical computations and array operations.

TensorFlow/Keras: Employed for constructing and training deep learning models, specifically for implementing LSTM and NHITS architectures.

Scikit-learn: Used for model performance evaluation and hyperparameter tuning.

During the model training process, the Adam optimizer was implemented, with the mean squared error (MSE) selected as the loss function. In the hyperparameter optimization phase, grid search was employed to identify the optimal combinations of hyperparameters. The specific optimization parameters are detailed in Table 1.

Table 1. Description of optimization iteration parameters

Parameter	Description
hidden_size	Dimension size of the LSTM hidden layer
n_blocks	Number of NHITS model blocks
n_hidden	Dimension of the hidden layer in each NHITS block
attention_size	Number of heads in the attention mechanism
learning_rate	Learning rate of the Adam optimizer

All models underwent hyperparameter optimization through grid search, using mean squared error (MSE) as the primary performance evaluation metric. Additionally, mean absolute error (MAE) and the coefficient of determination (R^2) were considered as auxiliary evaluation metrics. This systematic approach enabled the identification of optimal model configurations suitable for various time scales, thereby enhancing overall prediction performance.

3.4 Experiment Design and Results

3.4.1 Dataset Introduction

As illustrated in Figure 4, the study focuses on four monitoring sites in Zhanjiang City, which are critical for analyzing the ecological status and environmental changes of local water bodies. The water quality data collected from each site provides valuable insights into these dynamics. The scale of the dataset from each monitoring site is summarized in Table 2 below, highlighting the number of data points collected across multiple water quality parameters, enabling a dynamic analysis of changes over time.

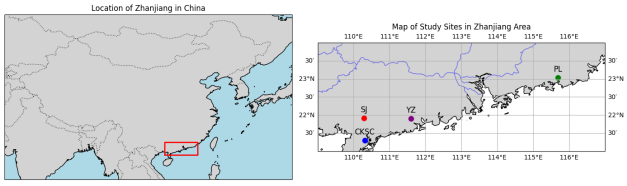

Figure 4. Monitoring site selection for surface water in Zhanjiang city

Figure 5 shows fluctuations in water quality parameters across the four monitoring sites, including water temperature, pH, dissolved oxygen, conductivity, and turbidity. Water temperature exhibits similar periodic fluctuations (10°C to 35°C) across all sites, likely influenced by environmental or seasonal factors. pH values are stable, mostly between 6 and 9, with SJ showing more variability compared to the more stable YZ site. Dissolved oxygen fluctuates significantly at SJ and PL, with occasional negative values, possibly due to sensor anomalies or unique conditions, while CKSC and YZ remain more stable. Conductivity shows dramatic changes, especially at CKSC and YZ, indicating short-term increases in salts or pollutants. Turbidity fluctuates widely, with extreme spikes at SJ and YZ, likely linked to rainfall, pollution, or sediment disturbances, while CKSC and PL are more stable with occasional peaks.

Table 2. Scale of each dataset

Monitoring site	Number of data points
SJ	3108
CKSC	2963
PL	3096
YZ	2947

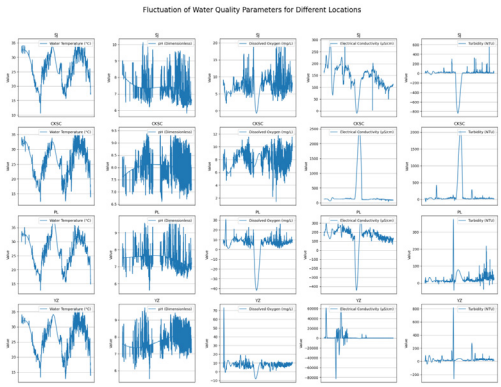

Figure 5. Fluctuations of water quality parameters for different locations

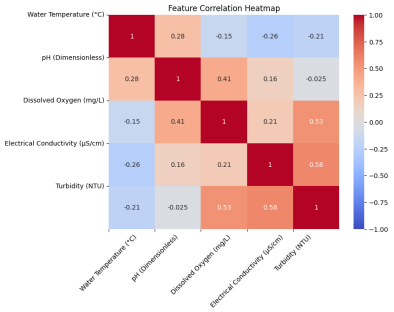

Figure 6. Heatmap of feature correlation analysis

Figure 6 presents correlation analysis results for the five parameters. Water temperature shows a weak positive correlation with pH (0.28), while dissolved oxygen has a weak negative correlation with temperature (-0.15). A moderate positive correlation between dissolved oxygen and pH (0.41) suggests that higher pH levels are linked to increased oxygen concentrations. Conductivity is positively correlated with both turbidity (0.58) and dissolved oxygen (0.21), indicating that higher conductivity corresponds with increased turbidity and oxygen, likely due to higher pollutants or suspended solids. Similarly, turbidity is positively correlated with dissolved oxygen (0.53), suggesting that increased turbidity may elevate organic content or pollutants, affecting oxygen levels.

These correlations highlight the complex interactions among water quality parameters, reinforcing the importance of considering them together in predictive models. The relationships between conductivity, turbidity, and dissolved oxygen are particularly significant, as they suggest that factors like pollution or sediment disturbances can influence water quality and prediction accuracy. Overall, while the parameters across sites show similar trends, their fluctuation characteristics differ, likely due to regional environmental conditions, seasonal changes, or unforeseen events. The data provide crucial insights for improving water quality monitoring and management strategies in the region.

3.4.2 Model Evaluation Metrics

Time series prediction models are evaluated using several metrics, including Mean Absolute Error (MAE), Mean Squared Error (MSE), and Coefficient of Determination (R^2) [20]. These metrics are calculated using the formulas in Equations (22), (23), and (24).

1. MAE measures the average absolute error between predicted and actual values, with a range of [0, +∞). It is less sensitive to outliers than MSE, providing a better reflection of average prediction error. Smaller MAE values indicate higher accuracy.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \tag{22}$$

2. MSE evaluates the average squared difference between predicted and actual values, also in the range of [0, +∞). Due to its squaring effect, MSE is more sensitive to larger errors and outliers. Smaller values indicate better performance.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 \tag{23}$$

3. R^2 assesses the goodness of fit, showing how well the predicted values match the actual values. Its range is (-∞, 1]. Higher R^2 values indicate a better fit, with 1 meaning perfect prediction and 0 meaning no better than predicting the mean. A negative R^2 suggests worse performance than a naive prediction.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} \tag{24}$$

y_i represents the actual value, $\hat{y}_i$ is the predicted value,

n is the number of samples, and $\bar{y}$ is the mean of the actual values.

These metrics collectively provide a comprehensive evaluation of the model's performance, guiding future optimization and improvement efforts.

3.4.3 Model Performance Evaluation

In this experiment, the cleaned dataset was split into 80% training and 20% testing sets. The LSTM+NHITS model was compared with traditional machine learning models (RF, XGBoost) and deep learning models (RNN, GRU, LSTM) across five water quality indicators. The hyperparameters tuned include LSTM hidden layer dimension (hidden_size), with values {32, 64, 128}, where 128 was chosen to balance efficiency and long-term dependency capture. The number of NHITS blocks (n_blocks) was tested between 1 and 5, with 2 blocks selected for efficient feature extraction. NHITS hidden layer dimension (n_hidden) was tested with values {64, 128, 256}, and 256 was chosen for optimal feature capture. The attention mechanism dimension (attention_size) was tested with values {32, 64, 128}, with 128 selected for effective emphasis on relevant features. The learning rate (learning_rate) was tested with values {0.0001, 0.001, 0.01}, and 0.001 was chosen for stable learning.

After cross-validation, the optimal hyperparameters were: hidden_size = 128, n_blocks = 2, n_hidden = 256, attention_size = 128, and learning_rate = 0.001. These settings provided a balance between performance and stability, with the lowest MSE indicating the model's robustness and efficiency in forecasting water quality time series data.

Figure 7 illustrates the attention mechanism weight distribution across 12 time steps, showing the model's ability to prioritize crucial periods in water quality predictions. The attention mechanism dynamically adjusts weights, focusing on significant fluctuations and trends, improving the model's accuracy and robustness. Figure 8 shows the training loss of the LSTM+NHITS model, demonstrating rapid convergence. Figure 9 compares predicted and actual values for the SJ site, showing that the model effectively captures short-term fluctuations and long-term trends in water quality indicators.

Performance results across ten models (see Table 3, Table 4, and Table 5) show that the LSTM+NHITS model significantly outperforms others in 4-hour predictions, with R^2 values 4% to 10% higher, and reductions in MAE (47%) and MSE (30% to 50%) [20]. The model's performance remains strong for 12-hour and 1-day forecasts, demonstrating its ability to handle both short-term and long-term variations. Figure 10 further illustrates the performance differences, highlighting the superiority of LSTM+NHITS in capturing temporal features.

The LSTM+NHITS model's success is due to LSTM's ability to capture long-term dependencies and NHITS's multi-head attention mechanism, which enhances focus on significant features. Traditional models like RF and XGBoost, though effective with large datasets, struggle with complex temporal patterns [21-22]. The LSTM+NHITS model's ability to capture both local and global features of time series data gives it a clear advantage. For 12-hour and 1-day forecasts, the LSTM+NHITS model showed a significant reduction in MAE compared to Autoformer, Informer, RF, and XGBoost, further validating its accuracy [23-24].

In conclusion, the LSTM+NHITS model proves to be a powerful tool for water quality prediction, offering significant improvements in forecasting accuracy and robustness, with broad applicability in environmental monitoring and future water quality management.

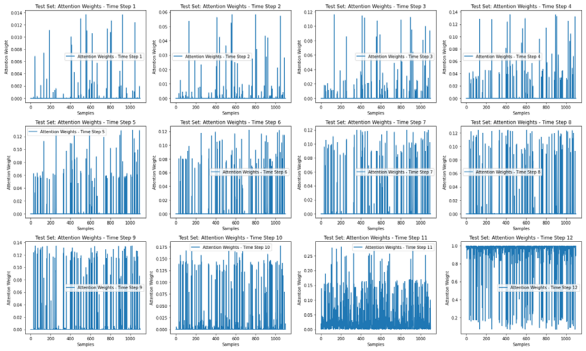

Figure 7. Attention mechanism weight distribution

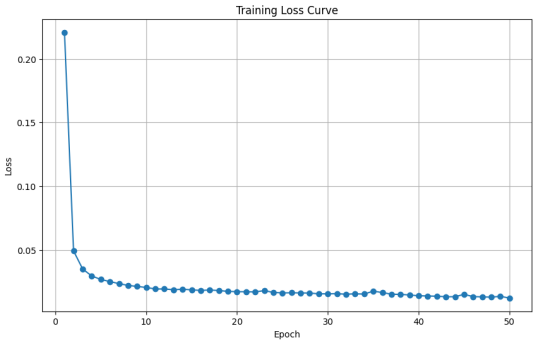

Figure 8. Training LOSS of the LSTM+NHITS model

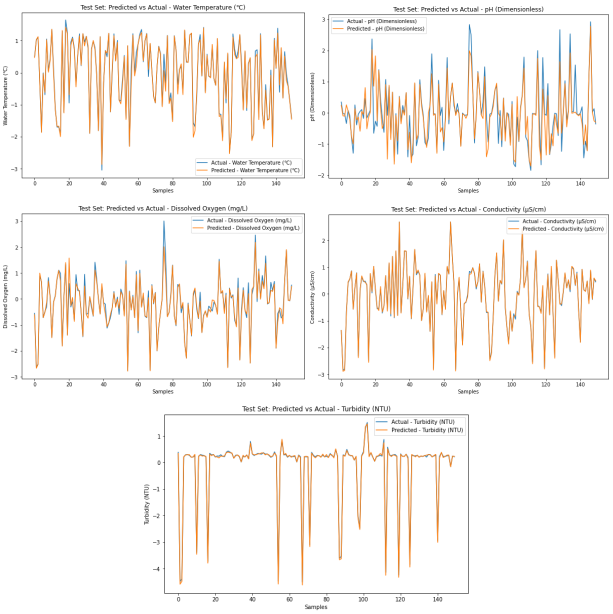

Figure 9. 4-Hour prediction curve and actual curve for SJ by the LSTM+NHITS model

Table 3. Evaluation of 4-hour prediction models

DataSet	Models	LSTM +NHITS	LSTM	NHITS	Autoformer
SJ	MAE	0.0542	0.1055	0.1090	0.1550
	MSE	0.0200	0.0586	0.0609	0.0986
	R2	0.9804	0.9386	0.9362	0.8968
CKSC	MAE	0.0534	0.1136	0.1219	0.2038
	MSE	0.0183	0.0714	0.0719	0.1913
	R2	0.9812	0.9309	0.9305	0.8151
PL	MAE	0.0549	0.0985	0.1037	0.1410
	MSE	0.0208	0.0555	0.0537	0.0882
	R2	0.9792	0.9428	0.9449	0.9107
YZ	MAE	0.0587	0.0973	0.1093	0.1725
	MSE	0.0208	0.0557	0.0610	0.1336
	R2	0.9795	0.9464	0.9413	0.8714

Informer	PatchTST	GRU	RNN	RF	XGboost
0.1692	0.1081	0.1099	0.1132	0.1021	0.0980
0.1148	0.0590	0.0672	0.0605	0.0605	0.0614
0.8798	0.9382	0.9296	0.9366	0.9367	0.9356
0.2229	0.1299	0.1155	0.1359	0.1159	0.1033
0.1971	0.0753	0.0674	0.0774	0.0815	0.0699
0.8093	0.9272	0.9347	0.9251	0.9213	0.9324
0.1585	0.1034	0.0968	0.1005	0.0922	0.0885
0.0989	0.0566	0.0549	0.0529	0.0540	0.0513
0.9008	0.9418	0.9434	0.9456	0.9445	0.9472
0.2033	0.1083	0.0975	0.1151	0.0989	0.0911
0.1565	0.0615	0.0574	0.0597	0.0652	0.0595
0.8498	0.9409	0.9448	0.9426	0.9373	0.9428

Table 4. Evaluation of 12-hour prediction models

DataSet	Models	LSTM +NHITS	LSTM	NHITS	Autoformer
SJ	MAE	0.0556	0.1115	0.1338	0.1898
	MSE	0.0181	0.0634	0.0749	0.1360
	R2	0.9809	0.9382	0.9270	0.8678
CKSC	MAE	0.0553	0.1079	0.1209	0.2240
	MSE	0.0193	0.0595	0.0629	0.2147
	R2	0.9806	0.9400	0.9366	0.7837
PL	MAE	0.0521	0.1002	0.1060	0.1648
	MSE	0.0174	0.0605	0.0596	0.1114
	R2	0.9822	0.9408	0.9417	0.8951
YZ	MAE	0.0635	0.0972	0.1090	0.2159
	MSE	0.0212	0.0532	0.0578	0.1776
	R2	0.9789	0.9480	0.9436	0.8290

Informer	PatchTST	GRU	RNN	RF	XGboost
0.2957	0.1296	0.1018	0.1146	0.1055	0.1002
0.2220	0.0705	0.0595	0.0607	0.0674	0.0626
0.7841	0.9313	0.9420	0.9409	0.9343	0.9390
0.2754	0.1290	0.1177	0.1181	0.1130	0.0995
0.2495	0.0670	0.0652	0.0583	0.0715	0.0587
0.7484	0.9326	0.9343	0.9413	0.9281	0.9409
0.2449	0.1210	0.0917	0.1042	0.0925	0.0912
0.1970	0.0619	0.0552	0.0608	0.0570	0.0591
0.8193	0.9398	0.9460	0.9406	0.9443	0.9423
0.2881	0.1360	0.0937	0.1145	0.0958	0.0930
0.2846	0.0682	0.0545	0.0580	0.0613	0.0569
0.7348	0.9337	0.9468	0.9433	0.9402	0.9447

Table 5. Evaluation of 1-day prediction models

DataSet	Models	LSTM +NHITS	LSTM	NHITS	Autoformer
SJ	MAE	0.0617	0.1058	0.1512	-
	MSE	0.0210	0.0608	0.0878	-
	R2	0.9800	0.9367	0.9084	-
CKSC	MAE	0.0561	0.1180	0.1405	-
	MSE	0.0178	0.0720	0.0811	-
	R2	0.9825	0.9323	0.9237	-
PL	MAE	0.0682	0.0954	0.1071	-
	MSE	0.0207	0.0559	0.0558	-
	R2	0.9776	0.9380	0.9383	-
YZ	MAE	0.0840	0.1015	0.1185	-
	MSE	0.0292	0.0654	0.0687	-
	R2	0.9675	0.9319	0.9281	-

Informer	PatchTST	GRU	RNN	RF	XGboost
0.3541	0.1414	0.0982	0.1211	0.1004	-
0.2802	0.0730	0.0587	0.0691	0.0608	-
0.7093	0.9243	0.9388	0.9282	0.9367	-
0.3187	0.1382	0.1165	0.1228	0.1202	-
0.2930	0.0837	0.0633	0.0661	0.0834	-
0.7242	0.9214	0.9405	0.9379	0.9218	-
0.2895	0.1422	0.0955	0.1008	0.0908	-
0.1966	0.0714	0.0559	0.0521	0.0532	-
0.7871	0.9212	0.9379	0.9422	0.9411	-
0.3366	0.1166	0.1021	0.1117	0.0977	-
0.2913	0.0705	0.0725	0.0632	0.0607	-
0.6823	0.9265	0.9252	0.9344	0.9359	-

Table 6. Results of ablation experiments

Experimental setting	MAE	MSE	R^2
Original model (All features)	0.0542	0.0200	0.9804
Only rolling average features, removing lag features	0.0802	0.0343	0.9663
Only lag features, removing rolling mean features	0.0566	0.0209	0.9795
Removing normalization step, using raw data directly	2.9770	41.7523	0.5869
Removing attention mechanism	0.0789	0.0248	0.9755
Removing random seed	0.0569	0.0202	0.9802
Using SGD optimizer	0.0675	0.0255	0.9737
Using RAdam optimizer	0.0548	0.0183	0.9820
Using gradient clipping and learning rate scheduling	0.0602	0.0190	0.9813

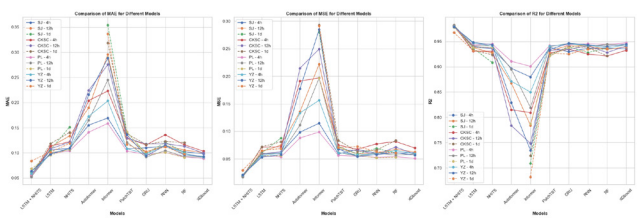

Figure 10. Comparison of model performance metrics

3.4.4 Ablation Experiment

To gain a deeper understanding of the contributions of different features and structures to water quality prediction performance, ablation experiments were conducted. These experiments involved systematically removing or adjusting certain features and components to assess their specific impact on model performance. The results of the experiments are presented in Table 6.

The original model, incorporating all features, performed best with an MAE of 0.0542, MSE of 0.0200, and R^2 of 0.9804. Using only rolling average features led to a significant decline in performance, with an MAE of 0.0802, highlighting the importance of lag features in capturing temporal dependencies. A model using only lag features had an MAE of 0.0566, underscoring the complementary nature of these features. Omitting the normalization step drastically reduced performance, with an MAE of 2.9770, while removing the attention mechanism increased the MAE to 0.0789, indicating its importance in focusing on significant features. Removing the random seed caused minor fluctuations (MAE = 0.0569), suggesting stability across different initializations. Using the SGD optimizer resulted in an MAE of 0.0675, indicating the advantage of the RAdam optimizer for this task.

The RAdam optimizer had a slightly higher MAE (0.0548), showing minimal impact on performance, but it improved stability and efficiency during training, with MSE and R^2 values consistent with the original model (MSE = 0.0200, R^2 = 0.9804). Gradient Clipping and Learning Rate Scheduling slightly increased the MAE to 0.0602, indicating minimal effect on overall performance (MSE = 0.0190, R^2 = 0.9813).

Despite some deviations, deep learning models performed variably in simulating water quality fluctuations. Research indicates that combining multiple models enhances accuracy [11]. Therefore, this study uses LSTM as a base for an ensemble approach, improving generalization and reducing overfitting risks. The LSTM+NHITS model reduced RMSE by an average of 12.0% (range: 5%–29.6%) compared to individual models, with R^2 increasing by 2% to 3%. For the 4-hour prediction, the MAE was 0.0542, reducing LSTM and NHITS MAEs by 48.6% and 50.3%, respectively. In 12-hour and 1-day predictions, the MAE was 0.0556 and 0.0617, confirming stability in short-term forecasting.

The LSTM+NHITS model excels in both accuracy and stability, with a negligible gap between training and testing R^2 values, demonstrating superior generalization. This ensemble approach offers robust support for water quality predictions, with broad applicability in environmental monitoring. Grid search optimization, due to its simplicity and clarity, consistently outperformed random search algorithms in terms of MAE, MSE, and R^2 metrics [25]. Table 7 shows the consistency of grid search across datasets, further validating its efficacy in hyperparameter optimization.

Table 7. 4-Hour prediction grid search vs. random search

DataSet	Models	Grid-LSTM-NHITS	LSTM-NHITS	Grid-LSTM	LSTM	Grid-NHITS	NHITS
SJ	MAE	0.0556	0.0561	0.1055	0.1086	0.1090	0.1144
	MSE	0.0181	0.0198	0.0586	0.0586	0.0609	0.0630
	R2	0.9809	0.9806	0.9386	0.9386	0.9362	0.9341
CKSC	MAE	0.0553	0.0607	0.1136	0.1172	0.1219	0.1177
	MSE	0.0193	0.0205	0.0714	0.0766	0.0719	0.0684
	R2	0.9806	0.9790	0.9309	0.9259	0.9305	0.9338
PL	MAE	0.0521	0.0501	0.0985	0.1042	0.1037	0.0957
	MSE	0.0174	0.0201	0.0555	0.0596	0.0537	0.0515
	R2	0.9822	0.9799	0.9428	0.9386	0.9449	0.9470
YZ	MAE	0.0635	0.0590	0.0973	0.0992	0.1093	0.1156
	MSE	0.0212	0.0222	0.0557	0.0628	0.0610	0.0601
	R2	0.9789	0.9782	0.9464	0.9396	0.9413	0.9422

4 Discussion

Local climate conditions, such as temperature, rainfall, and humidity, can directly impact water quality by influencing factors like water temperature, nutrient levels, and the likelihood of algal blooms. Industrial activities, particularly those related to water usage and emissions, can introduce pollutants that degrade water quality. Additionally, land use changes, such as urbanization and agricultural expansion, can alter water runoff patterns, introducing sediments and chemicals into water bodies, which further affects water quality.

Integrating these external variables into the LSTM-NHITS model has the potential to enhance its predictive capability by offering a more complete picture of the environmental factors influencing water quality. For example, including climate data could allow the model to better account for seasonal variations in water quality. On the other hand, incorporating industrial and land use data could improve predictions during extreme events like industrial discharges or heavy rainfall.

However, adding these variables is not without its challenges. Collecting accurate and timely data on external factors can be difficult, especially in regions with underdeveloped monitoring systems. Moreover, integrating these variables requires sophisticated feature engineering and preprocessing to avoid introducing noise or bias into the model. Adding more variables also increases

the complexity of the model, which may lead to longer training times and potentially lower prediction accuracy if not carefully calibrated. Therefore, selecting and validating the most impactful external variables is critical to ensuring the success of the model.

Looking forward, future research could explore integrating these external variables to develop a more robust forecasting system. This could involve utilizing publicly available data, such as satellite imagery or datasets from government agencies, to better capture the factors influencing water quality. Furthermore, the development of real-time monitoring systems that continuously collect and integrate these variables could allow for more accurate predictions, helping decision-makers take timely actions, such as issuing alerts during pollution events or extreme weather. Finally, transfer learning techniques could be considered to apply the model to other regions, taking into account the unique environmental factors of different areas and expanding the model's applicability beyond its current geographic scope.

5 Conclusion

This study introduces an improved LSTM-NHITS model for water quality monitoring and prediction, addressing challenges related to accuracy and applicability. The model outperformed traditional machine learning techniques (such as Random Forest and XGBoost) and deep learning methods (including RNN, GRU, and LSTM), showing better performance across various prediction time frames. Experimental results from monitoring sites in Zhanjiang City demonstrate the model's capability to predict water quality fluctuations accurately, achieving reductions in MAE and MSE of up to 47%. For the 4-hour forecast, the LSTM-NHITS model achieved R^2 values that were 4% to 10% higher than other models and reduced MSE by 30% to 50%. The model also performed well for 12-hour and 1-day forecasts, effectively managing both short-term fluctuations and long-term trends.

By leveraging LSTM's ability to capture long-term dependencies and NHITS's attention-based feature selection, the model accurately predicts sudden changes in water quality, making it a valuable tool for real-time prediction and environmental management. Future research could look into combining multiple deep learning models, incorporating external factors like climate change and land use, and developing real-time monitoring systems. Additionally, transfer learning could extend the model's applicability to different regions and water bodies.

While the results are promising, certain challenges to validity should be considered. Issues with data quality, such as missing or noisy data, could impact model performance, highlighting the need for robust data preprocessing techniques. Moreover, the exclusion of external variables—such as local climate conditions or land use changes—limits the model's ability to generalize. The model's reliance solely on water quality parameters may reduce its adaptability to regions with different environmental characteristics and pollution sources. Additionally, the long-term prediction capability of the model remains uncertain, requiring further research to evaluate its performance over extended periods.

In conclusion, the LSTM-NHITS model provides a powerful tool for water quality prediction, with the potential for further improvement through the inclusion of external variables and real-time data.

6 Acknowledgement

This work was partially supported by the Fund of Southern Marine Science and Engineering Guangdong Laboratory (Zhanjiang) (No. ZJW-2023-04); the National Key Research and Development Program of China (No. 2022YFD2401304); the Scientific Research Project of the Education Department of Guangdong Province (No. 2024ZDZX4060, 2022GCZX001, 2023ZDZX4012); Special Fund Project for Talent Development Strategy of Guangdong Province (No. 2024R1003); Zhanjiang Science and Technology Plan Project (Grant No. 2024B01067); and the program for scientific research start-up funds of Guangdong Ocean University (Grant No. 060302102302).

References

[1] S. Khurana, G. Sharma, B. Sharma, Hybrid Machine Learning Model for Load Prediction in Cloud Environment, *International Journal of Performability Engineering*, Vol. 19, No. 8, pp. 507-515, August, 2023. https://doi.org/10.23940/ijpe.23.08.p3.507515

[2] H. Zuo, X. Gou, X. Wang, M. Zhang, A combined model for water quality prediction based on VMD-TCN-ARIMA optimized by WSWOA, *Water*, Vol. 15, No. 24, Article No. 4227, December, 2023. https://doi.org/10.3390/w15244227

[3] A. Nsabimana, P. Li, Y. Wang, S. K. Alam, Variation and multi-time series prediction of total hardness in groundwater of the Guanzhong Plain (China) using grey Markov model, *Environmental Monitoring and Assessment*, Vol. 194, No. 12, Article No. 899, December, 2022. https://doi.org/10.1007/s10661-022-10585-9

[4] S. Zhang, A. H. Omar, A. S. Hashim, T. Alam, H. A. E. W. Khalifa, M. A. Elkotb, Enhancing waste management and prediction of water quality in the sustainable urban environment using optimized algorithm of least square support vector machine and deep learning techniques, *Urban Climate*, Vol. 58, Article No. 102095, November, 2024. https://doi.org/10.1016/j.uclim.2024.102095

[5] V. Diaz, W. E. Wong, Z. Chen, Enhancing Deception Detection with Exclusive Visual Features using Deep Learning, *International Journal of Performability Engineering*, Vol. 19, No. 8, pp. 547-558, August, 2023. https://doi.org/10.23940/ijpe.23.08.p7.547558

[6] J. Nyyssölä, M. Mäntylä, Event-level Anomaly Detection on Software logs: Role of Algorithm, Threshold, and Window Size, *2024 IEEE 24th International Conference on Software Quality, Reliability and Security*, Cambridge, United Kingdom, 2024, pp. 649-656. https://doi.org/10.1109/QRS62785.2024.00070

[7] Y. Sun, Z. Lian, S. Zhang, Z. Wang, T. Duan, A Novel Explainable Method based on Grad-CAM for Network

Intrusion Detection, *2024 IEEE 24th International Conference on Software Quality, Reliability and Security*, Cambridge, United Kingdom, 2024, pp. 400-406. https://doi.org/10.1109/QRS62785.2024.00046

[8] S. Rajagopal, S. S. Ganesh, A. Karthick, T. Sampradeepraj, Environmental water quality prediction based on COOT-CSO-LSTM deep learning, *Environmental Science and Pollution Research*, Vol. 31, No. 42, pp. 54525-54533, September, 2024. https://doi.org/10.1007/s11356-024-34750-4

[9] M. Gauch, F. Kratzert, D. Klotz, G. Nearing, J. Lin, S. Hochreiter, Rainfall–runoff prediction at multiple timescales with a single Long Short-Term Memory network, *Hydrology and Earth System Sciences*, Vol. 25, No. 4, pp. 2045-2062, April, 2021. https://doi.org/10.5194/hess-25-2045-2021

[10] K. Zhang, X. Wang, T. Liu, W. Wei, F. Zhang, M. Huang, H. Liu, Enhancing water quality prediction with advanced machine learning techniques: An extreme gradient boosting model based on long short-term memory and autoencoder, *Journal of Hydrology*, Vol. 644, Article No. 132115, November, 2024. https://doi.org/10.1016/j.jhydrol.2024.132115

[11] K. Wang, L. Liu, X. Ben, D. Jin, Y. Zhu, F. Wang, Hybrid deep learning based prediction for water quality of plain watershed, *Environmental Research*, Vol. 262, Article No. 119911, December, 2024. https://doi.org/10.1016/j.envres.2024.119911

[12] D. R. Gandh, V. P. Harigovindan, K. P. R. Abdul Haq, A. Bhide, Attention-driven LSTM and GRU deep learning techniques for precise water quality prediction in smart aquaculture, *Aquaculture International*, Vol. 32, No. 6, pp. 8455-8478, December, 2024. https://doi.org/10.1007/s10499-024-01574-5

[13] A. Boufekane, M. Meddi, D. Maizi, G. Busico, Performance of artificial intelligence model (LSTM model) for estimating and predicting water quality index for irrigation purposes in order to improve agricultural production, *Environmental Monitoring and Assessment*, Vol. 196, No. 11, Article No. 1049, November, 2024. https://doi.org/10.1007/s10661-024-13211-y

[14] C. Challu, K. G. Olivares, B. N. Oreshkin, F. Garza Ramirez, M. Mergenthaler-Canseco, A. Dubrawski, NHITS: Neural Hierarchical Interpolation for Time Series Forecasting, *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence*, Vol. 37, No. 6, pp. 6989-6997, June, 2023. https://doi.org/10.1609/aaai.v37i6.25854

[15] M. Pazo, S. Gerassis, M. Araújo, I. M. Antunes, X. Rigueira, Enhancing water quality prediction for fluctuating missing data scenarios: A dynamic Bayesian network-based processing system to monitor cyanobacteria proliferation, *The Science of the Total Environment*, Vol. 927, Article No. 172340, June, 2024. https://doi.org/10.1016/j.scitotenv.2024.172340

[16] S. Purwani, R. A. Hidayana, V. P. Balqis, Sukono, A Comparison of Newton's Divided Differences Interpolation and a Cubic Spline in Predicting the Poverty Rate of West Java, *Engineering Letters*, Vol. 31, No. 4, pp. 1786-1791, November, 2023.

[17] J. Liu, B. Jin, J. Yang, L. Xu, Sea surface temperature prediction using a cubic B-spline interpolation and spatiotemporal attention mechanism, *Remote Sensing Letters*, Vol. 12, No. 5, pp. 478-487, March, 2021. https://doi.org/10.1080/2150704X.2021.1897182

[18] D. Antanasijević, V. Pocajt, A. Perić-Grujić, M. Ristić, Modeling of dissolved oxygen in the Danube River using artificial neural networks and Monte Carlo Simulation uncertainty analysis, *Journal of Hydrology*, Vol. 519, pp. 1895-1907, November, 2014. https://doi.org/10.1016/j.jhydrol.2014.10.009

[19] Y. Zheng, J. Wang, Y. Kondratenko, J. Wu, Research progress in surface water quality monitoring based on remote sensing technology, *International Journal of Remote Sensing*, Vol. 45, No. 7, pp. 2337-2373, March, 2024. https://doi.org/10.1080/01431161.2024.2327086

[20] N. A. Zaini, A. N. Ahmed, L. W. Ean, M. F. Chow, M. A. Malek, Forecasting of fine particulate matter based on LSTM and optimization algorithm, *Journal of Cleaner Production*, Vol. 427, Article No. 139233, November, 2023. https://doi.org/10.1016/j.jclepro.2023.139233

[21] Q. V. Ly, V. H. Truong, B. Ji, X. C. Nguyen, K. H. Cho, H. H. Ngo, Z. Zhang, Exploring potential machine learning application based on big data for prediction of wastewater quality from different full-scale wastewater treatment plants, *The Science of the Total Environment*, Vol. 832, Article No. 154930, August, 2022. https://doi.org/10.1016/j.scitotenv.2022.154930

[22] J. Chen, J. Wan, G. Ye, Y. Wang, Prediction and optimization of wastewater treatment process effluent chemical oxygen demand and energy consumption based on typical ensemble learning models, *Bioresource Technology*, Vol. 411, Article No. 131362, November, 2024. https://doi.org/10.1016/j.biortech.2024.131362

[23] H. Wu, J. Xu, J. Wang, M. Long, Autoformer: Decomposition Transformers with auto-correlation for Long-Term Series Forecasting, *NIPS'21: Proceedings of the 35th International Conference on Neural Information Processing System*, Online, 2021, pp. 22419-22430.

[24] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, W. Zhang, Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting, Proceedings of the *AAAI Conference on Artificial Intelligence*, Vol. 35, No. 12, pp. 11106-11115, May, 2021. https://doi.org/10.1609/aaai.v35i12.17325

[25] Z. Cömert, A. Sbrollini, F. Demircan, L. Burattini, Computerized otoscopy image-based artificial intelligence model utilizing deep features provided by vision transformer, grid search optimization, and support vector machine for otitis media diagnosis, *Neural Computing and Applications*, Vol. 36, no. 36, pp. 23113-23129, December, 2024. https://doi.org/10.1007/s00521-024-10457-y

Biographies

Xiaohong Peng is a professor of computer science at Guangdong Ocean University. She graduated from the Computer Science program at South China Normal University with a master's degree. Her main research areas are intelligent control and intelligent information processing.

Tianrong Zhong graduated with a master's degree in Computer Science and Technology from Guangdong Ocean University. His research focuses on intelligent networks and systems.

Renyou Yang graduated from the University of Strathclyde. He is currently a researcher at the Southern Marine Science and Engineering Guangdong Laboratory. His main research areas are intelligent systems.

Zhao Li received the MS and PhD degrees in computer science and technology from Wuhan University. He is a professor in Computer Science and Technology, Guangdong Ocean University. His research focuses on image recognition, big data processing, and software defects prediction.