Journal of Geoscience and Environment Protection, 2025, 13(3), 347-363
https://www.scirp.org/journal/gep
ISSN Online: 2327-4344
ISSN Print: 2327-4336

Reliable Water Quality Prediction Using Bayesian Multi-Scale Convolutional Attention Network

Xiaolin Guo

Key Laboratory of Three Gorges Reservoir Region's Eco-Environment, Ministry of Education, Chongqing University, Chongqing, China
Email: cqugxl@126.com

How to cite this paper: Guo, X. L. (2025). Reliable Water Quality Prediction Using Bayesian Multi-Scale Convolutional Attention Network. *Journal of Geoscience and Environment Protection, 13,* 347-363.
https://doi.org/10.4236/gep.2025.133019

Received: January 23, 2025
Accepted: March 28, 2025
Published: March 31, 2025

Abstract

With the rapid development of industrialization and urbanization, the issue of water quality deterioration has become increasingly severe. Accurately assessing water quality is crucial for environmental protection and public health. Traditional water quality testing methods rely on sampling and laboratory analysis, which are costly and inefficient. In recent years, artificial intelligence (AI) based techniques have gained attention in research on water quality prediction because of their effectiveness and advanced capabilities. However, the black-box nature of AI model makes it difficult to quantify the reliability of their predictions, limiting their practical application. To address this issue, this paper proposes a Bayesian multi-scale convolutional attention network for water quality prediction. This method extracts high-level features affecting water quality through a multi-scale convolutional network and combines a self-attention mechanism and gated feature fusion approach to enhance the representation of key features and effectively integrate information. At the same time, Bayesian inference is used to generate prediction confidence intervals, providing a reliable assessment for the results. To the best of our knowledge, no research has yet combined Bayesian methods with deep learning for water quality prediction. Experimental results on the Kaggle water quality dataset demonstrate that the proposed method not only performs excellently in prediction accuracy but also effectively quantifies prediction uncertainty, providing scientific support for water quality assessment.

Keywords

Uncertainty Quantification, Water Quality Prediction, Feature Fusion

1. Introduction

The safety of drinking water quality is one of the basic needs of human survival and an important basis for maintaining the ecological balance of the earth. However, the swift progress of industrialization and urbanization has exposed Earth's surface water to substantial human interference and eutrophication, causing a decline in water quality, ecosystem damage, and the loss of vital ecological functions (Zhang et al., 2024). This pollution of the water environment has become a pressing issue, threatening the safety of drinking water sources and the stability of regional water supplies (Lv et al., 2024). Therefore, water quality prediction for drinking water is of great significance for improving the health of the global population and maintaining the health of aquatic ecosystems (Zhou, Wang, Xiao, Wang, & Sun, 2018).

The traditional water quality detection method involves extracting water samples from water bodies, transporting them to a laboratory for a series of chemical and physical measurements, and then analyzing the results to draw conclusions (Dogo, Nwulu, Twala, & Aigbavboa, 2019). Although this method can provide reliable water quality data, it is costly, inefficient, and requires substantial expertise. Fortunately, the rapid development of artificial intelligence (AI) technologies has provided new approaches for water quality prediction (Yin et al., 2025). Currently, the use of AI techniques for water quality prediction has become a hot topic in the research field (Baena-Navarro, Carriazo-Regino, Torres-Hoyos, & Pinedo-López, 2025).

The AI-based water quality prediction methods can be divided into shallow machine learning based methods and deep learning based methods. Initially, the shallow machine learning methods are used for water quality prediction. Compared with the method of manual sampling and sending to the laboratory for analysis, the method based on shallow machine learning greatly improves the efficiency of water quality prediction. Li et al. (Li, Lu, Wu, Zhang, & Chen, 2022) compare the performance of methods such as Radial Basis Function Neural Network (RBFNN) and Support Vector Machine (SVM) in water quality prediction and find that SVM exhibits the best performance. Xu et al. (Xu et al., 2021) develop a water quality prediction framework based on Random Forest. Lu et al. (Lu & Ma, 2020) add a denoising module to the methods based on Extreme Gradient Boosting (XGBoost) and Random Forest (RF) to predict the water quality of the Tualatin River. Although the shallow machine learning based water quality prediction methods can overcome the high cost and low efficiency of traditional methods, they are unable to fully capture the deep level features in the data that affect water quality, which limits the improvement of their predictive performance. The development of deep learning provides a pathway to overcome this challenge. Deep learning methods have the advantage of capturing the deep intrinsic connections of data, which makes the use of deep learning methods for water quality prediction mainstream. Yu et al. (Yu & Xiao, 2024) use an Long short-term memory (LSTM) based method to predict the water quality of the

Haihe River Basin. Alfwzan et al. (Alfwzan, Selim, Althobaiti, & Hussin, 2023) propose a deep learning-based Bi-LSTM model to predict the variables affecting groundwater quality. Zheng et al. (Zheng, Ding, Weng, & Wang, 2024) use a neural network based on a dual-attention mechanism for water quality prediction, which improves prediction accuracy compared to the LSTM method. Hu et al. (Hu, Lyu, Wang, Zhou, & Fang, 2023) combine the attention mechanism with Temporal Convolutional Networks (TCN), effectively improving the model's performance in water quality prediction. Although the above works overcome the limitations of machine learning-based methods, most of the aforementioned methods overlook the deep relationships between the features affecting water quality, which limits their ability to achieve better water quality prediction performance. He et al. (He, Wu, Huang, Kang, & Gui, 2022) extract multi-scale temporal features of water quality, improving prediction performance. However, they do not consider the multi-scale complex relationships between the features affecting water quality. For example, in some cases, it may be necessary to focus on a single feature (e.g., pH value), while in other cases, the relationships between multiple features (e.g., hardness, dissolved oxygen) need to be jointly considered to assess water quality.

Although deep learning based water quality prediction methods offer the advantages of high efficiency and low cost, their lack of interpretability poses a challenge in providing guidance to decision-makers due to low reliability. Liu et al. (M. Y. Liu et al., 2023) apply Bayesian methods to water quality prediction tasks, providing confidence intervals for the prediction results. However, the relatively weak inference capability of Bayesian methods may lead to certain limitations when the aforementioned works face more complex water quality prediction tasks.

To address the aforementioned challenges, we combine Bayesian methods with deep learning approaches, while also focusing on the multi-scale relationships of the features that influence water quality. Specifically, we propose an end-to-end water quality prediction method based on a Bayesian multi-scale convolutional attention network. In our proposed method, we first introduce a feature capture module based on multi-scale CNN, which captures the relationships between features from both global-to-local and local-to-global perspectives. Next, we propose a dynamic fusion module based on the attention mechanism, which adjusts the focus on the features influencing water quality and dynamically integrates the extracted global and local features. Additionally, we employ a Bayesian approach to obtain the confidence intervals for water quality predictions, thereby providing a reliability assessment of the results. This enables decision makers to make more informed and well-founded decisions in water quality management and policy formulation. The main contributions of this article are summarized as follows.

1) We propose an end-to-end water quality prediction method based on a Bayesian multi-scale CNN attention network. And to the best of our knowledge, no research has yet combined Bayesian methods with deep learning for water quality prediction. Experiments on the Kaggle water quality prediction dataset

demonstrate the effectiveness of our proposed method.

2) We use a Bayesian approach to obtain the confidence intervals for water quality predictions, which can quantify the uncertainty of the predictions, thereby providing decision makers with reliable assessments of the forecasted results.

3) We propose a multi-scale CNN feature capturing method that can capture the relationships between features from both global-to-local and local-to-global perspectives, thereby enhancing the model's performance in water quality prediction.

4) We propose a dynamic feature fusion method based on an attention network. This method can adjust the focus on key factors influencing water quality and dynamically integrate multi-scale features through a gating mechanism. The remainder of this article is organized as follows. Section II defines the research problem addressed in this paper. Section III details the structure and theory of the proposed water quality prediction method. Section IV introduces the experimental setup and comparative experiments. Section V concludes the paper.

2. Problem Definition

In this study, the water quality prediction is conducted by assessing the health status of water using multiple indicators.

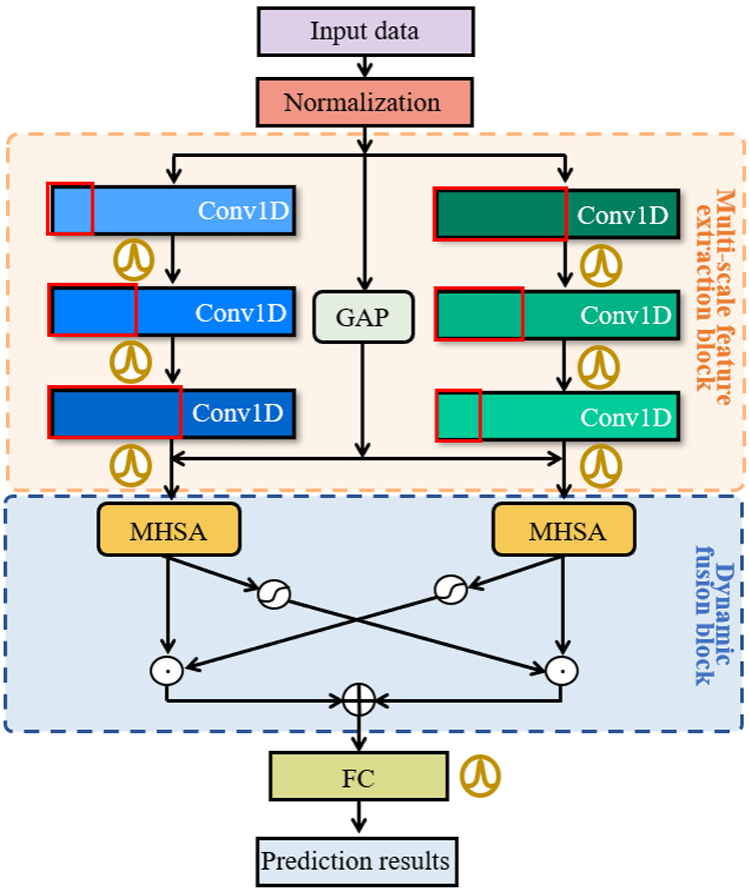

Figure 1. Structure of the proposed method.

Suppose a water sample has n indicators, denoted as $x=[x_1, x_2, ..., x_n]$ and the water quality status is represented as $y \in \{1, ..., C\}$ where C indicates the number of water quality levels. In this study, $C=2$. The water quality prediction task in this study can be formulated as:

$$x \xrightarrow{f(\cdot)} y \tag{1}$$

The task of this study is to learn the mapping function $f(\cdot)$ from water features to water quality health states, enabling water quality prediction.

3. Methodology

This section provides a detailed description of the proposed water quality prediction network, as shown in **Figure 1**. We will first describe the overall network architecture, followed by a detailed introduction to the proposed multi-scale feature extraction block and dynamic fusion block. Finally, we will explain the process of obtaining the confidence intervals using Bayesian inference.

3.1. Overall Network Architecture

As shown **Figure 1**, we propose an end-to-end water quality prediction network. First, we normalize the input data and then feed it into the multi-scale feature extraction block we designed. The multi-scale feature extraction block leverages Bayesian convolution to extract key features influencing water quality from both global-to-local and local-to-global perspectives. Next, the extracted multi-scale features are input into the dynamic fusion block based on multi-head attention (MHSA) mechanism. This block adjusts the network's focus on features and dynamically integrates features from different scales using a gating mechanism. Finally, the fused features are passed into the Bayesian fully connected layers to obtain the prediction results. To quantify the uncertainty of the predictions, we use Bayesian inference to compute the confidence intervals of the predicted results.

3.2. Multi-Scale Feature Extraction Block

In the process of water quality assessment, a single indicator is sometimes sufficient for making a judgment, such as when the pH value is too high or too low. However, in many cases, a single indicator is not enough to comprehensively evaluate water quality, and a joint analysis of multiple indicators is necessary. Most existing studies often overlook the local and global relationships within water quality features (Mei, Li, Zhang, Li, & Song, 2022; Yang et al., 2023). To address this issue, we design the multi-scale feature extraction block, with the core idea of progressively extracting and capturing key features affecting water quality from both global-to-local and local-to-global perspectives.

Specifically, as shown in the multi-scale feature extraction block in **Figure 1**, the normalized data is input into the multi-scale feature extraction block and divided into three branches. The left branch consists of three layers of Bayesian convolutional networks, with progressively larger kernel sizes. This design allows the

network's receptive field to gradually expand, shifting the focus from capturing local features to capturing global features. For the first layer of Bayesian convolution, given the input X, assume that the weights W_{L1} and bias b_{L1} follow a Gaussian distribution, which can be formulated as

$$W_{L1} \sim \eta(\mu_{W_{L1}}, \sigma^2_{W_{L1}}), \tag{2}$$

and

$$b_{L1} \sim \eta(\mu_{b_{L1}}, \sigma^2_{b_{L1}}). \tag{3}$$

The output of the first Bayesian convolution is

$$Z_{L1} = W'_{L1} \cdot X[i : i + K_{L1}] + b'_{L1} \tag{4}$$

where W'_{L1} represents the sampled values of the weights in the first layer of Bayesian convolution, b'_{L1} represents the sampled values of the biases in the first layer of Bayesian convolution, i represents the starting position of the convolution operation, and K_{L1} represents the size of the convolution window during each operation. Similarly, we can obtain the output Z_{L3} of the third layer of Bayesian convolution on the left. Similar to the three layers of Bayesian convolution on the left, the network on the right also consists of three layers of Bayesian convolutions. However, the kernel sizes in the right-side network decrease from large to small, starting with global feature extraction and gradually focusing on local features. Using the same computation method, the output of the third layer on the right can be denoted as Z_{R3}. Unlike the left and right branches, the middle branch is designed as a residual branch. Given the input X, after passing through the global average pooling (GAP) layer, the resulting feature is added to Z_{R3} and Z_{L3}.

This design, by introducing residual connections, can enhance the multi-scale feature extraction block's perception of global information, thereby improving both the training speed and model stability.

3.3. Dynamic Fusion Block

The features influencing water quality have varying levels of importance, and there are also interdependencies among these features. Inspired by Wu et al. (Wu et al., 2019), we design the dynamic fusion block, as shown in **Figure 1**. The core idea of our design is to assign different importance weights to multi-scale features using MHSA, and then dynamically fuse features from different scales through gating mechanisms.

Specifically, for the two outputs of the multi-scale feature extraction block, we first input them into the MHSA. The structure of MHSA is shown in **Figure 2**. Given the input X, it is passed through the fully connected layers to obtain the queries (Q), keys (K), and values (V).

$$Q^m = \sigma_Q(W_Q \cdot X + b_Q), \tag{5}$$

$$K^m = \sigma_K(W_K \cdot X + b_K), \tag{6}$$

$$V^m = \sigma_V(W_V \cdot X + b_V), \tag{7}$$

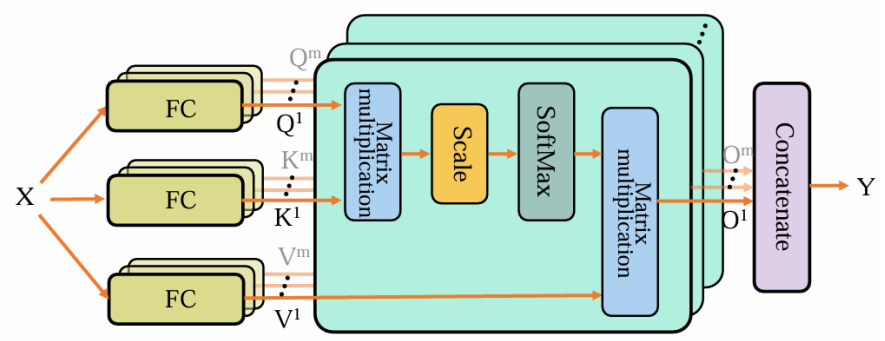

Figure 2. Structure of the MHSA.

where Q^m, K^m and V^m are the m-th heads of Q, K and V respectively, σ_Q, σ_K and σ_V are activation functions, W_Q, W_K and W_V are weights, b_Q, b_K and b_V are biases. Then, the Q and K are multiplied and passed through a softmax function to obtain the attention scores. The final output of MHSA (Y) is obtained by multiplying the attention scores with the V and then concatenating them, which can be formulated as

$$O^{(m)} = soft\max(\frac{Q^m K^{(m)T}}{\sqrt{d}})V^m, m = 1,...,M \,, \tag{8}$$

$$Y = Concat(O^{(1)},...,O^{(M)}) \,, \tag{9}$$

where d is the dimension of each head.

Let the outputs of the MHSA on the left and right be denoted as Y_L and Y_R, respectively. The output of the dynamic fusion block is

$$F = sigmoid(Y_R) \odot Y_L + sigmoid(Y_L) \odot Y_R \tag{10}$$

where $\odot$ denotes element-wise multiplication. And it is worth mentioning that in the left part, the sigmoid activation function maps Y_R (or Y_L) to the range of 0 to 1. The result is then multiplied by Y_L (or Y_R), which amplifies the important features in Y_L (or Y_R) and reduces the less important ones, thus highlighting the key features that influence water quality.

3.4. Bayesian Inference

Our proposed method uses Bayesian inference to obtain the confidence intervals of the predicted results. In our proposed method, both the convolutional networks and fully connected layers are implemented in a Bayesian form. Unlike traditional convolutional networks and fully connected layers, during training, the model learns not only the mean of the parameters but also their standard deviation. This enables the model to quantify the uncertainty of the parameters.

The main idea of using Bayesian methods to obtain the confidence intervals of the predicted results is as follows: by sampling multiple different parameter values from the probability distribution of the parameters and performing multiple forward passes. Each forward pass will yield different prediction results, and these results can be used to compute the mean and standard deviation of the predictions.

Finally, based on the mean and standard deviation, we can construct the confidence intervals of the predictions, thereby quantifying the uncertainty of the predictions. Specifically, we sample the model's prediction output for N times. The mean and standard deviation of these predictions can be expressed as

$$\mu_S = \frac{1}{N}\sum_{i=1}^{N} S_i , \tag{11}$$

and

$$\sigma_S = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(S_i - \mu_S)^2} , \tag{12}$$

where S_i is the prediction result obtained from the i-th sampling, μ_S is the mean of the predicted results, σ_S is the standard deviation of the predicted results. If the confidence level is set to 95%, the 95% confidence interval (CI) can be expressed as

$$CI = [\hat{S}_{mean} - 1.96 \cdot \hat{\sigma}_S, \hat{S}_{mean} + 1.96 \cdot \hat{\sigma}_S] . \tag{13}$$

4. Experiment

In this section, we will first introduce the dataset used, followed by the experimental setup and evaluation metrics. Subsequently, we will conduct comparative experiments, sensitivity analyses, and ablation studies to validate the superiority, robustness, and necessity of each component of the proposed method.

4.1. Dataset Description

This study uses the dataset provided by Kaggle (Chakravarthy et al., 2023). The dataset contains 3276 records, with each record consisting of 9 features, including pH, Hardness, Solids, Chloramines, Sulfate, Conductivity, Organic Carbon, Trihalomethanes, and Turbidity, refer to **Table 1**. The target label is Potability, which indicates whether the water is safe to drink, with a value of 0 meaning unsafe to drink and 1 meaning safe to drink.

Table 1. Feature description.

Feature	Description
pH	Water's acidity or alkalinity
Hardness	Concentration of calcium and magnesium ions
Solids	Total dissolved and suspended solids in water
Chloramines	Disinfection by-products formed during chlorination
Sulfate	Concentration of sulfate salts in water
Conductivity	Water's ability to conduct electricity
Organic carbon	Amount of organic carbon in water
Trihalomethanes	Disinfection by-products that may pose health risks
Turbidity	Clarity of water, indicating suspended particles

4.2. Experimental Setup

1) Data normalization: Since the value ranges of the features in the dataset differ significantly, some features may have an outsized influence on the model during training, while others may have less impact. To address this issue and ensure that all features are compared on the same scale, thereby improving the model's training efficiency and accuracy, this study normalizes the input features. The process of normalizing the feature Xi can be formulated as

$$X_i^{'} = \frac{X_i - X_{\min,i}}{X_{\max,i} - X_{\min,i}}, \tag{14}$$

where $X_{\min,i}$ and $X_{\max,i}$ are the minimum and maximum values of the i-th feature, respectively.

2) Missing value imputation: As shown in **Figure 3**, missing values exist in the features Trihalomethanes, Sulfate, and ph. Among them, Sulfate has the most missing values, with 781 missing entries, followed by pH, with 491 missing entries.

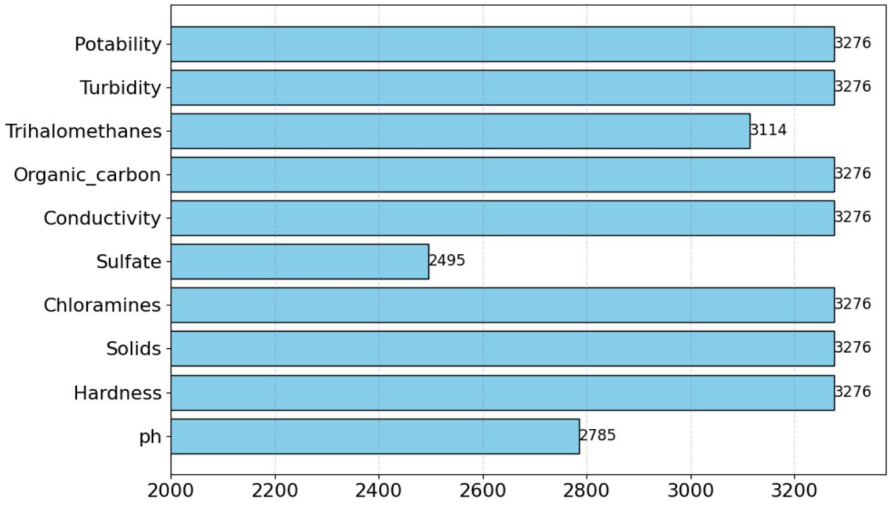

Figure 3. Number of data entries for features and labels.

In this paper, we used the Iterative Imputation method (J. C. Liu, Gelman, Hill, Su, & Kropko, 2014) to fill in the missing values. The main process is as follows: first, the missing values are initialized with the mean of the feature; then, other features without missing values are used as independent variables to train a linear regression model to predict the missing values of the current feature. After 10 iterations, the imputation results are gradually optimized, and the missing feature values are ultimately estimated. In this study, we use 80% of the data for training and the remaining data for testing.

4.3. Evaluation Metrics

In this paper, we select Accuracy, Precision, Recall, and F1-Score as evaluation metrics (Arslan et al., 2025). Each metric measures the model's performance from different perspectives. The metric Accuracy can be formulated as

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \tag{15}$$

where TP is the number of correctly predicted positive samples, TN is the number of correctly predicted negative samples, FP is the number of samples incorrectly predicted as positive and FN is the number of samples incorrectly predicted as negative. The metric precision evaluates the ratio of actual positive instances among those predicted as positive.

$$Precision = \frac{TP}{TP + FP}. \tag{16}$$

The Recall metric emphasizes the model's capability to identify positive instances

$$Recall = \frac{TP}{TP + FN}. \tag{17}$$

And the metric F1-Score is the harmonic mean of Precision and Recall, combining both metrics into a single value, which provides a balance between Precision and Recall

$$F1\text{-}Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}. \tag{18}$$

4.4. Model Performance

Figure 4 shows the performance of the proposed method across four evaluation metrics. The model achieves a precision of 0.96, recall of 0.98, accuracy of 0.98, and an F1-score of 0.97. This demonstrates that the proposed method performs well in predicting both positive and negative samples.

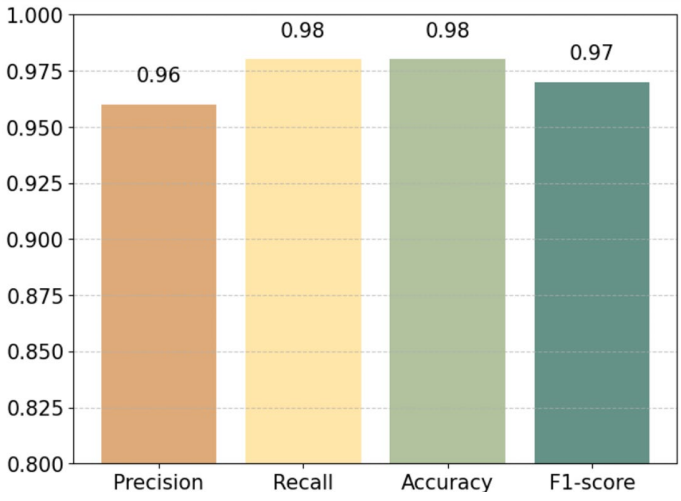

Figure 4. Model performance.

Figure 5 shows the predicted probability range of the proposed method under a 95% confidence interval for 20 test samples (where the proposed method correctly predicted all 20 test samples). It can be observed that the proposed method

has different prediction ranges for different test samples, which can help decision-makers make more reliable and robust decisions.

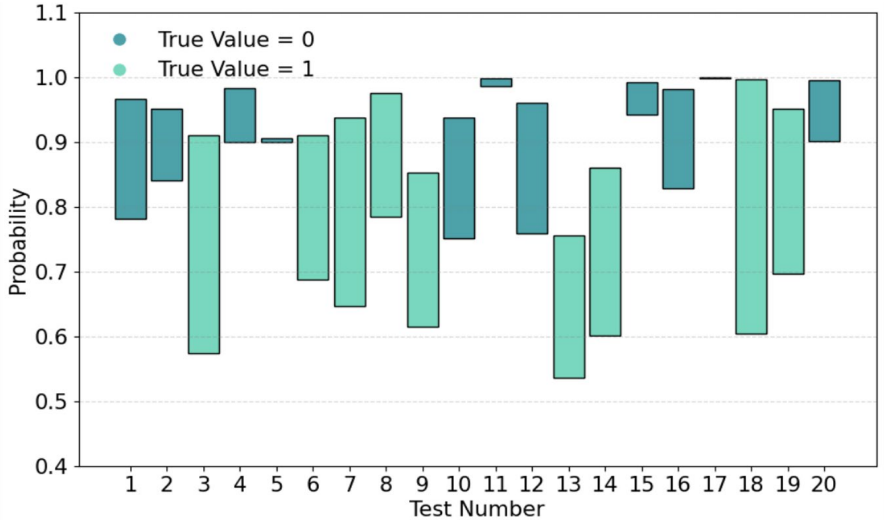

Figure 5. 95% confidence intervals for the predicted probabilities of 20 test samples.

4.5. Comparative Experiments

To demonstrate the superiority of the proposed method, we compare its prediction performance with classic machine learning methods such as SVM, K-Nearest Neighbors(KNN), DecisionTree(DT), RF, and XGBoost, as well as state-of-the art deep learning works including ANN (Rustam et al., 2022) and LSTM (P. Liu, Wang, Sangaiah, Xie, & Yin, 2019).

Table 2. Results of the comparison experiments.

Methods	Precision	Recall	Accuracy	F1-score
SVM	0.31	0.50	0.63	0.39
KNN	0.60	0.59	0.64	0.59
DT	0.59	0.59	0.61	0.59
RF	0.66	0.61	0.68	0.61
XGBoost	0.61	0.60	0.65	0.60
ANN	0.82	0.52	0.77	0.64
LSTM	0.70	0.32	0.68	0.44
Proposed	0.96	0.98	0.98	0.97

As shown in **Table 2** and **Figure 6**, our method outperforms all others across the four evaluation metrics, demonstrating the superiority of the proposed method. Comparing the proposed method with ANN and LSTM, it can be observed that, although the latter are also deep networks, they are unable to capture the relationships between water quality features, resulting in suboptimal performance in water quality prediction. In contrast to shallow machine learning methods,

the proposed approach outperforms the shallow machine learning methods by a large margin. This is because the shallow machine learning methods fail to recognize the intricate connections between water quality features and water quality states.

To further compare the dependence of the above methods on the data, we reduce the training set to 20% and 10%. **Table 3** and **Figure 7** show the model performance when only 20% of the data is used for training. As can be seen from the table, with the reduction in the training data, most methods experience varying degrees of performance degradation. The proposed method still demonstrates the best performance. The prediction performance of ANN and LSTM decreases the most, as deep learning methods depend significantly on vast quantities of training data. In contrast, the proposed method mitigates the dependence on data by capturing the complex relationships between water quality features and adjusting the focus, enabling it to uncover the essential features that influence water quality states. **Table 4** and **Figure 8** show the comparison of model performance

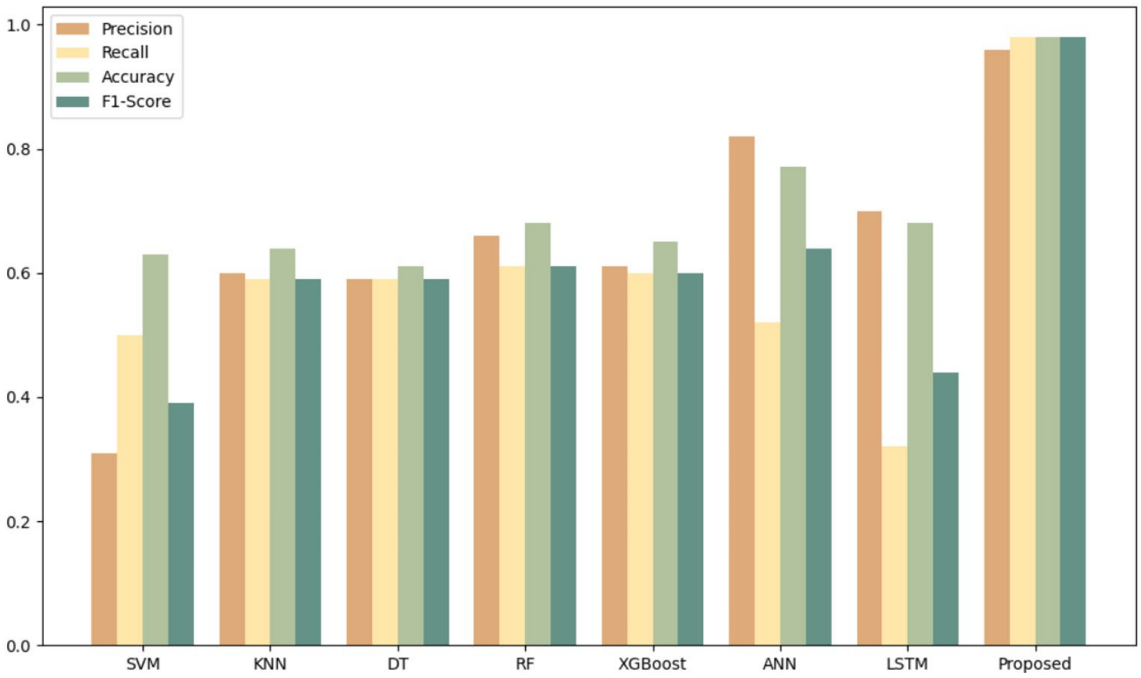

Figure 6. Results of the comparison experiments.

Table 3. Comparison of the results using 20% of the total dataset for training.

Methods	Precision	Recall	Accuracy	F1-score
Scheme A	0.79	0.73	0.81	0.76
Scheme B	0.76	0.83	0.85	0.79
Scheme C	0.64	0.71	0.70	0.67
Scheme D	0.69	0.71	0.74	0.70
Proposed	**0.81**	**0.90**	**0.88**	**0.86**

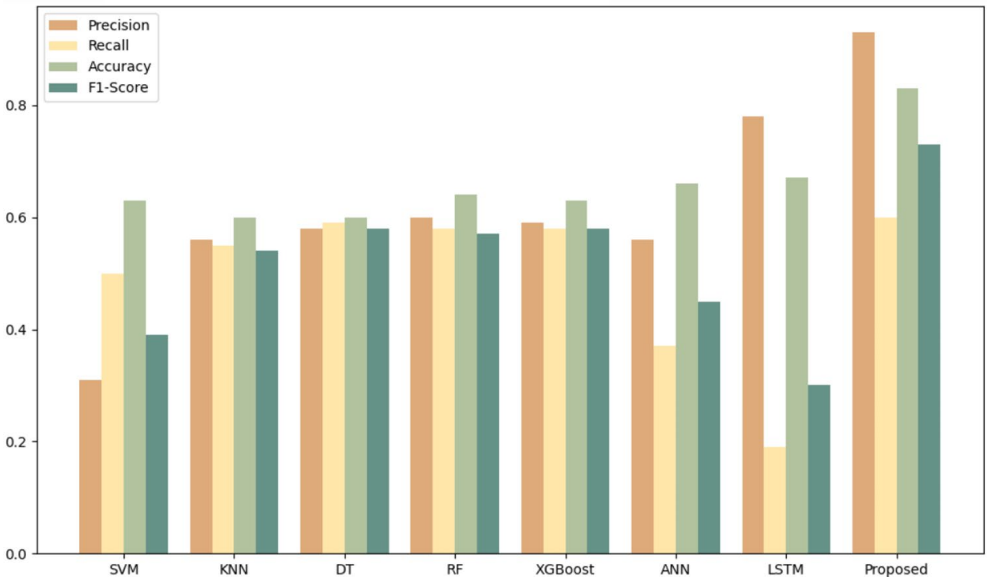

Figure 7. Comparison of the results using 20% of the total dataset for training.

Table 4. Comparison of the results using 10% of the total dataset for training.

Methods	Precision	Recall	Accuracy	F1-score
SVM	0.31	0.50	0.63	0.39
KNN	0.55	0.55	0.59	0.55
DT	0.56	0.56	0.58	0.56
RF	0.62	0.59	0.65	0.59
XGBoost	0.57	0.56	0.61	0.56
ANN	0.53	0.36	0.64	0.43
LSTM	0.54	0.42	0.65	0.48
Proposed	**0.81**	**0.90**	**0.88**	**0.86**

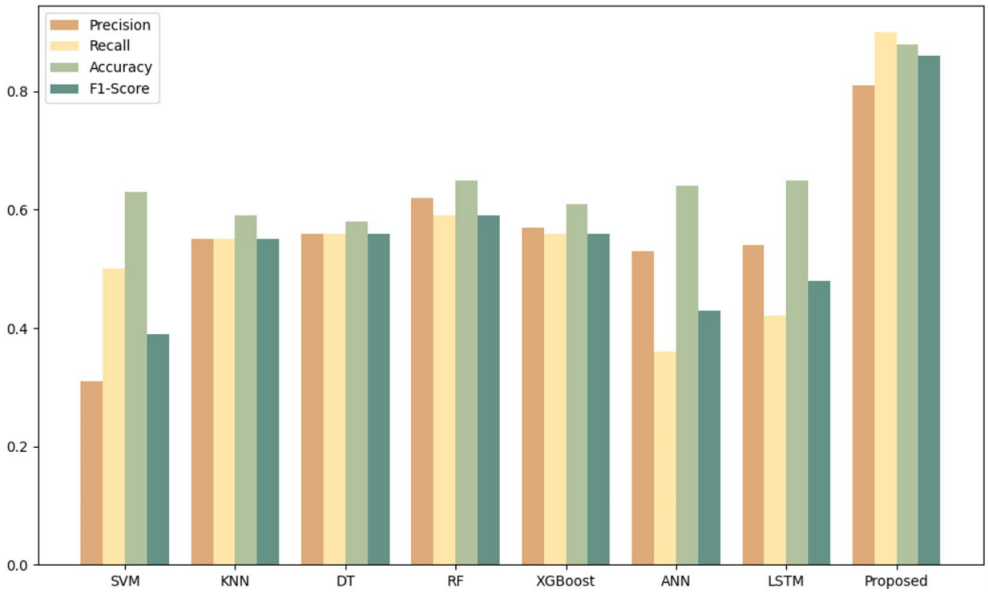

Figure 8. Comparison of the results using 10% of the total dataset for training.

when the training set is reduced to 10% of the total dataset. It can be observed that the proposed method still outperforms the others, further proving its superiority.

4.6. Sensitivity Analysis

To investigate the sensitivity of the proposed method to parameters, we compared the performance of the method with Bayesian convolution layers set to 2 and 4. As shown in **Figure 9**, overall, the proposed method exhibits a certain level of robustness to changes in the number of Bayesian convolution layers. When the number of convolution layers is 2, the model performance decreases the most compared to when the number of convolution layers is 4. This is because the model is not able to fully learn both local and global features with too few convolution layers. When the number of convolution layers is 4, the model's performance also decreases. This is because the model tends to overfit with a larger number of layers.

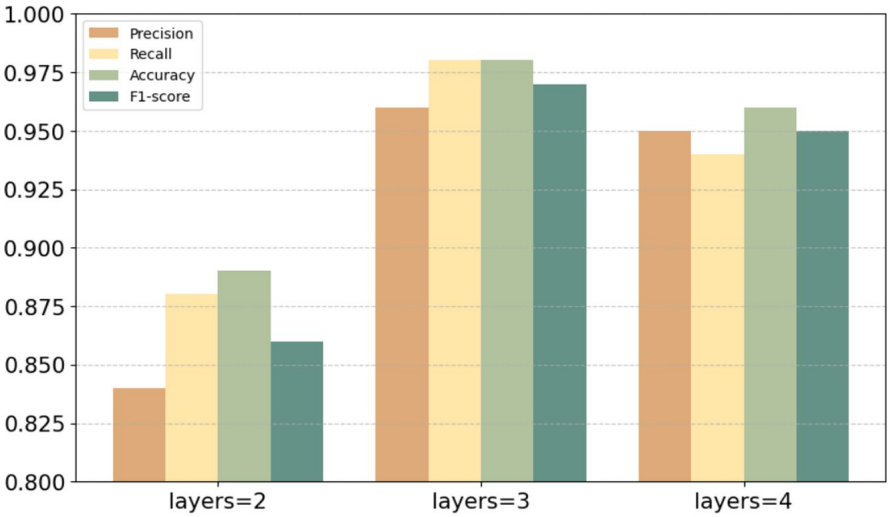

Figure 9. Model performance with different numbers of Bayesian convolutional Layers.

4.7. Ablation Experiments

To demonstrate the necessity of each element in the proposed model, we designed ablation experiments. Specifically, we sequentially remove the left Bayesian convolutional layer in the multi-scale feature extraction block (A), the right Bayesian convolutional layer in the multi-scale feature extraction block (B), the entire multi-scale feature extraction block (C), and the dynamic fusion block (D).

Table 5 presents the results of the ablation experiments. As can be seen from the table, the proposed method achieves the best performance, demonstrating the necessity of each component in the proposed method. Comparing scheme A and B, it can be observed that removing the left convolutional network leads to a more significant performance decline, which indicates that capturing the relationships between water quality features from local to global is more effective. Comparing scheme C and D, it can be seen that removing the multi-scale feature extraction

block results in a greater performance drop. This is because, without the multi-scale feature extraction block, the dynamic fusion block is unable to effectively capture the deep connections between water quality features and water quality states.

Table 5. Results of the ablation experiments.

Methods	Precision	Recall	Accuracy	F1-score
SVM	0.31	0.50	0.63	0.39
KNN	0.55	0.55	0.59	0.55
DT	0.56	0.56	0.58	0.56
RF	0.62	0.59	0.65	0.59
XGBoost	0.57	0.56	0.61	0.56
ANN	0.53	0.36	0.64	0.43
LSTM	0.54	0.42	0.65	0.48
Proposed	**0.81**	**0.90**	**0.88**	**0.86**

5. Conclusion

In this paper, we propose an end-to-end water quality prediction network using Bayesian multi-scale convolutional attention network. Our proposed method is capable of simultaneously capturing both global and local features that influence water quality and dynamically integrating them. More importantly, the proposed method provides the confidence intervals of the prediction results, offering decision-makers with more reliable and robust insights for water quality management. Our comparative experiments demonstrate the superiority of the proposed method. In future research, we will explore the temporal changes in water quality and investigate the time-series prediction of key water quality features.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

Alfwzan, W. F., Selim, M. M., Althobaiti, S., & Hussin, A. M. (2023). Application of Bi-LSTM Method for Groundwater Quality Assessment through Water Quality Indices. *Journal of Water Process Engineering, 53,* Article ID: 103889. https://doi.org/10.1016/j.jwpe.2023.103889

Arslan, M., Asad, M., Haider Khan, A., Iqbal, S., Nabeel Asghar, M., & Abdulrhman Alaulamie, A. (2025). Deep Image Synthesis, Analysis and Indexing Using Integrated CNN Architectures. *IEEE Access, 13,* 834-851. https://doi.org/10.1109/access.2024.3515455

Baena-Navarro, R., Carriazo-Regino, Y., Torres-Hoyos, F., & Pinedo-López, J. (2025). Intelligent Prediction and Continuous Monitoring of Water Quality in Aquaculture: Integration of Machine Learning and Internet of Things for Sustainable Management. *Water, 17,* Article No. 82. https://doi.org/10.3390/w17010082

Chakravarthy, S. R. S., Bharanidharan, N., Venkatesan, V. K., Abbas, M., Rajaguru, H.,

Mahesh, T. R. et al. (2023). Prediction of Water Quality Using Soft-Max-ELM Optimized Using Adaptive Crow-Search Algorithm. *IEEE Access, 11,* 140900-140913. https://doi.org/10.1109/access.2023.3339564

Dogo, E. M., Nwulu, N. I., Twala, B., & Aigbavboa, C. (2019). A Survey of Machine Learning Methods Applied to Anomaly Detection on Drinking-Water Quality Data. *Urban Water Journal, 16,* 235-248. https://doi.org/10.1080/1573062x.2019.1637002

He, M., Wu, S., Huang, B., Kang, C., & Gui, F. (2022). Prediction of Total Nitrogen and Phosphorus in Surface Water by Deep Learning Methods Based on Multi-Scale Feature Extraction. *Water, 14,* Article No. 1643. https://doi.org/10.3390/w14101643

Hu, Y., Lyu, L., Wang, N., Zhou, X., & Fang, M. (2023). Application of Hybrid Improved Temporal Convolution Network Model in Time Series Prediction of River Water Quality. *Scientific Reports, 13,* Article No. 11260. https://doi.org/10.1038/s41598-023-38465-3

Li, T., Lu, J., Wu, J., Zhang, Z., & Chen, L. (2022). Predicting Aquaculture Water Quality Using Machine Learning Approaches. *Water, 14,* Article No. 2836. https://doi.org/10.3390/w14182836

Liu, J., Gelman, A., Hill, J., Su, Y.-S., & Kropko, J. (2014). On the Stationary Distribution of Iterative Imputations. *Biometrika, 101,* 155-173. https://doi.org/10.1093/biomet/ast044

Liu, M., Hu, J., Huang, Y., He, J., Effiong, K., Tang, T. et al. (2023). Probabilistic Prediction of Algal Blooms from Basic Water Quality Parameters by Bayesian Scale-Mixture of Skew-Normal Model. *Environmental Research Letters, 18,* Article ID: 014034. https://doi.org/10.1088/1748-9326/acaf11

Liu, P., Wang, J., Sangaiah, A. K., Xie, Y., & Yin, X. (2019). Analysis and Prediction of Water Quality Using LSTM Deep Neural Networks in IoT Environment. *Sustainability, 11,* Article No. 2058. https://doi.org/10.3390/su11072058

Lu, H., & Ma, X. (2020). Hybrid Decision Tree-Based Machine Learning Models for Short-Term Water Quality Prediction. *Chemosphere, 249,* Article ID: 126169. https://doi.org/10.1016/j.chemosphere.2020.126169

Lv, J., Du, L., Lin, H., Wang, B., Yin, W., Song, Y. et al. (2024). Enhancing Effluent Quality Prediction in Wastewater Treatment Plants through the Integration of Factor Analysis and Machine Learning. *Bioresource Technology, 393,* Article ID: 130008. https://doi.org/10.1016/j.biortech.2023.130008

Mei, P., Li, M., Zhang, Q., Li, G., & song, L. (2022). Prediction Model of Drinking Water Source Quality with Potential Industrial-Agricultural Pollution Based on CNN-GRU-Attention. *Journal of Hydrology, 610,* Article ID: 127934. https://doi.org/10.1016/j.jhydrol.2022.127934

Rustam, F., Ishaq, A., Kokab, S. T., de la Torre Diez, I., Mazón, J. L. V., Rodríguez, C. L. et al. (2022). An Artificial Neural Network Model for Water Quality and Water Consumption Prediction. *Water, 14,* Article No. 3359. https://doi.org/10.3390/w14213359

Wu, Z. H., Pan, S. R., Long, G. D. et al. (2019). *Graph WaveNet for Deep Spatial-Temporal Graph Modeling.* https://arxiv.org/abs/1909.05257

Xu, J., Xu, Z., Kuang, J., Lin, C., Xiao, L., Huang, X. et al. (2021). An Alternative to Laboratory Testing: Random Forest-Based Water Quality Prediction Framework for Inland and Nearshore Water Bodies. *Water, 13,* Article No. 3262. https://doi.org/10.3390/w13223262

Yang, J., Jia, L., Guo, Z., Shen, Y., Li, X., Mou, Z. et al. (2023). Prediction and Control of Water Quality in Recirculating Aquaculture System Based on Hybrid Neural Network. *Engineering Applications of Artificial Intelligence, 121,* Article ID: 106002.

https://doi.org/10.1016/j.engappai.2023.106002

Yin, H., Chen, Y., Zhou, J., Xie, Y., Wei, Q., & Xu, Z. (2025). A Probabilistic Deep Learning Approach to Enhance the Prediction of Wastewater Treatment Plant Effluent Quality under Shocking Load Events. *Water Research X, 26,* Article ID: 100291. https://doi.org/10.1016/j.wroa.2024.100291

Yu, A., & Xiao, Q. (2024). A Water Quality Prediction Model Based on Long Short-Term Memory Networks and Optimization Algorithms. *IEEE Access, 12,* 175607-175615. https://doi.org/10.1109/access.2024.3487348

Zhang, Y., Deng, J., Zhou, Y., Zhang, Y., Qin, B., Song, C. et al. (2024). Drinking Water Safety Improvement and Future Challenge of Lakes and Reservoirs. *Science Bulletin, 69,* 3558-3570. https://doi.org/10.1016/j.scib.2024.06.018

Zheng, Z., Ding, H., Weng, Z., & Wang, L. (2024). Research on Out-of-Sample Prediction Method of Water Quality Parameters Based on Dual-Attention Mechanism. *Environmental Modelling & Software, 176,* Article ID: 106020. https://doi.org/10.1016/j.envsoft.2024.106020

Zhou, J., Wang, Y., Xiao, F., Wang, Y., & Sun, L. (2018). Water Quality Prediction Method Based on IGRA and LSTM. *Water, 10,* Article No. 1148. https://doi.org/10.3390/w10091148