An Improved Temporal Convolutional Network with Residual Multi-head Attention for Long-term Water Quality Prediction

Subarna Sapkota[1*], Subash Timalsina[2], Prajwal Rai[3], Dipendra Ghimire[4]
[1] Student, Nepal College of Information Technology, Pokhara University
(ORCID: 0009-0002-2535-8239)
[2]Student, Nepal College of Information Technology, Pokhara University
(ORCID: 0009-0007-6450-7811)
[3]Student, Kantipur City College, Purbanchal University
(ORCID: 0009-0002-0051-4450)
[4]Lecturer, Ara institute of Canterbury
(ORCID: 0000-0003-1336-7541)

Abstract

Accurate prediction of river water quality based on time-series data is essential for understanding variations and protecting river environments. However, forecasting accuracy is severely influenced by strong seasonality, nonlinearity, and periodicity. Traditional statistical methods suffer from low accuracy, high temporal complexity, and poor long-term prediction capability, particularly in dynamic and constantly changing surroundings in open waters. To address these deficiencies, a novel deep learning framework based on a Temporal Convolutional Network (TCN) integrated with a residual multi-head attention mechanism is presented to forecast dissolved oxygen (DO). Initially, the raw data of water quality parameters were smoothed by using Savitzky Golay (SG) filter, which is then decomposed into three components: trend, seasonal, and residual components. The trend and residual components are supplied with the enhanced TCN model for training and prediction. Dilated causal convolutions enable effective extraction of long-term temporal dependencies, while the multi-head attention mechanism improves the capability to capture both local and long-range dependencies in time series data, thereby enhancing feature representation and prediction accuracy. Moreover, when compared with deep learning models such as LSTM, GRU and traditional statistical approaches, the proposed model consistently outperformed all baseline methods with R^2 values of 0.9906 and 0.9651 for Short-term (1 day) and Long-term (7 days) respectively. The findings confirmed that the proposed model provides a robust and effective solution for complex river water quality prediction tasks and offers valuable support for sustainable water resource management and environmental protection.

Keywords: Water quality, TCN, Multi-head attention, Time series,

[*]Corresponding Author Email: Subarnasapkota1@gmail.com
Published: 30 March 2026
DOI: https://doi.org/10.70558/IJST.2026.v3.i1.241202

1. Introduction

1.1 Background

Water is integral to life for the reason that it ensures agricultural productivity, maintains environmental balance, and guarantees healthy drinking water (Haghiabi et al., 2018). Worldwide, water quality problems are growing worse; nevertheless, as a result of the social economy, industry, and technological development are occurring so quickly. anticipating water quality has become crucial for managing water resources and reducing environmental damage, especially in light of growing worldwide concerns about availability and pollution levels. Water quality has a significant impact on public health and the standard of living of locals along with functioning as a direct indicator of regional economic progress (Liang & Zhang, 2024). Predicting and evaluating water quality is essential for ensuring public and environmental health as well as for efficient and sustainable use of water resources (Li et al., 2017). (Hu et al., 2023) applied the TCN model by adding self attention mechanism to predict the river water quality. TCN is a type of neural network designed specifically for sequence modeling tasks, which can include time-series forecasting, language modeling, and more. TCNs leverage the principles of convolutional neural networks (CNNs) while adapting them to handle sequential data effectively by allowing parallel processing also capture long-range dependencies more effectively due to their ability to use dilated convolutions. The accuracy of the model is contributed by the self-attention mechanism through the capturing of internal correlations within the data. However, the model's ability to capture intricate and diverse relationships in long-term water quality data is limited by the use of a single attention head. Additionally, greater sensitivity to noise in water quality data is observed, as the single attention head may be overly influenced by noisy or outlier observations.

In order to solve the problems existing in the above models and improve the performance of long term water quality prediction, a TCN model integrated with multi head attention mechanism is proposed(Morandini et al., 2023). In this system, sampling and mean imputation are first used during the data preprocessing stage. Missing values in the dataset are replaced with the Forward Fill method. After that, the SG filter is used for smoothing data and decomposing these time series data into seasonal, trend and residual series. The decomposed trend series and residual series are input into the model, combining a multi-head attention mechanism and TCN, respectively, for training and prediction. The TCN model, which uses multiple residual blocks, dilated convolution, and causal convolution, due to this, TCN supports parallel computation; the individual weights in each layer can be updated simultaneously at each time step, greatly increasing the model's computational efficiency (Fu et al., 2021). TCN also has a larger memory capacity and does not have the gradient vanishing problem as the RNN model does. To enhance the TCN predictive power and address the issue of local information loss, the multi-head attention procedure with the TCN model is implemented, in which the attention model allows the model to capture complex patterns within the water quality data by focusing more on those features that are contributing more to the output.

1.2 Problem Statement

Due to factors including noise and variations in the seasons, current water quality prediction algorithms have struggled to properly represent the complex nature of river water quality data. (Hu et al., 2023) applied a TCN model with self-attention to predict the water quality parameters. However, this research has some limitations due to its single-head form, including difficulties in capturing multi-dimensional relationships among various factors influencing water quality, challenges in identifying intricate patterns in residuals, and increased sensitivity to noise in the data. As a result, the model performed poorly when forecasting at long-term periods. To address these limitations, the TCN model with a multi-head attention mechanism is presented. This focus distribution makes the model more immune to noise and better able to represent intricate fluctuations in water quality data. As a result of its ability to integrate and balance various variables in the data, the model can make predictions that are more accurate over longer time periods.

1.3 Research objectives

The main objective of this research is to enhance performance in long-term water quality prediction by integrating TCN with a multi-head attention mechanism.

1.4 Significance of the study

The proposed research on a water quality prediction system based on time series data is crucial for protecting public health, informing water management decisions, and ensuring regulatory compliance. Accurate predictions can prevent waterborne diseases, optimise resource allocation, and enhance pollution control, benefiting communities, water authorities, and environmental agencies. The study advances scientific understanding, fosters community engagement, and supports sustainable water management.

2. Literature Review

2.1. Related Theory

Over time, forecasting water quality has grown in importance. Many researchers have used information from a variety of sources, such as satellite images, weather data, river flow measurements, and chemical analysis of water samples, to study the quality of water. Most of the prior research developed statistical models connecting these attributes to water quality indicators, such as the water quality index (WQI), using decision trees, random forests, support vector regression (SVR), and linear regression. In order to estimate the long-term water quality, various techniques have been employed recently, including time series water data and TCN, BI-LSTM, and LSTM models.(Shams et al., 2023) applied machine learning models based on the grid search method for Water quality prediction. Various machine learning models, such as Random Forest, Extreme Gradient Boosting, and K-Nearest Neighbour regressors, were utilized for classification and regression tasks. The testing results demonstrated that the GB model using the grid search technique performed the most accurate results in terms of classification, with a 99.5% prediction accuracy for WQC values, while in regression tasks, the Multi-Layer Perceptron regressor model fared better than other models in terms of its performance and acquired the best results with R^2 of 99.8% when predicting WQI values.

2.2 Long Short-Term Memory (LSTM) Model:

Zhou et al. (2018) offered a technique that integrated long short-term memory (LSTM) and enhanced grey-relational analysis (GRA) to improve the model's robustness and its applicability(Zhou et al., 2018). Regardless of experimental results, the suggested approach improved on single-feature or non-sequential prediction methods in water quality prediction by fully employing the multiple-variable correlations and chronological order of water quality observations. However, it has some limitations, such as high time complexity for prediction and training the model and consumes more memory. A unique hybrid model that combined a Savitzky-Golay filter with an encoder-decoder neural network based on long short-term memory (LSTM) is proposed by Bi et al. (2021) for large-scale water quality prediction. The challenges faced by traditional linear models in handling the complex and noisy nature of time series data on water quality have been addressed by this method. While the LSTM network identifies nonlinear relationships in the data to improve the model's prediction power, the Savitzky-Golay filter is used to minimise noise. The proposed SE-LSTM model performed better in terms of prediction accuracy than a number of modern benchmarks, according to experimental results, suggesting that it could be a useful tool for efficient management of water resources. In Figure 1, this paper is primarily focused on the analysis of LSTM memory cells at time step t ∈ T. LSTM memory cell is composed of three separate gates, namely forget, input and output gates. The activation function in this model is represented by using the symbol σ. Additionally, all three gates ($f_t; i_t; o_t$) used the sigmoid function to alter the memory at t, whereas c_t indicates the state of the memory cell,h_t point out the output of an LSTM unit. $f_t; i_t$ and o_t are the corresponding calculation methods for input, output and forget at time t.

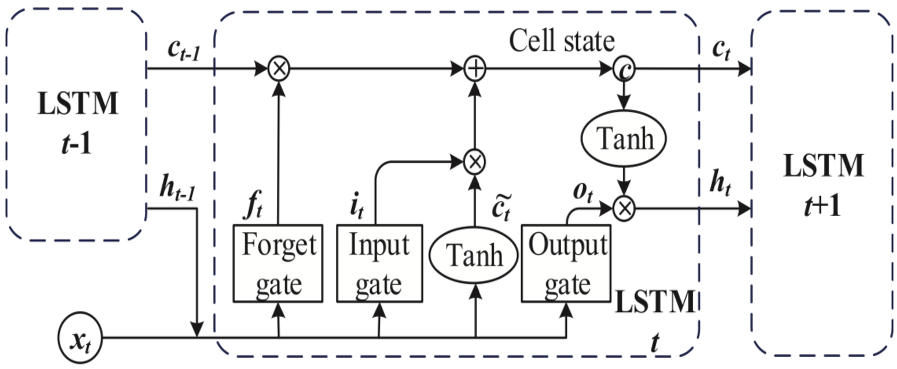

Figure 1: LSTM Units (Bi et al., 2021)

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \tag{1}$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \tag{2}$$

$$\widetilde{c_t} = tanh(W_f[h_{t-1}, x_t] + b_c) \tag{3}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \widetilde{c_t} \tag{4}$$

In the output gate, the sigmoid function specifies the dropped part of a state of a cell. The tanh function deals with the current cell state ct, and it is further multiplied by the result of the sigmoid function. Then, the following output at $t0_t$, isproduced

$$o_t = \sigma(W_0[h_{t-1}, x_t] + b_0) \quad (5)$$

$$h_t = o_t \odot \tanh(c_t) \quad (6)$$

Where $\odot$ denotes the operation of dot-wise product, the matrices$W_i, W_f \, and W_o$ and Wo denote the parameters of the input, forget and output gates, respectively. W_C stands for memory cell parameters. The symbols tanh(.) and $\sigma(.)$ indicate hyperbolic and sigmoid functions, respectively.

2.3 Temporal Convolutional Network (TCN) Model:

Temporal Convolutional Network(TCN) model used to predict the temperature of water, dissolved oxygen value and PH value (Fu et al. 2021). The TCN model reduced the likelihood of gradient explosion or disappearance during training by exhibiting a stable gradient. Reliable long-term water quality forecasts were produced by the model, which was very helpful in supporting aquaculture management decisions. In particular, this study concluded that the long-term prediction method of water quality parameters is capable of predicting tha water quality components for upcomming ten days with an prediction accuracy of 91.91%, an average decrease in training time of 64.92%, and an median decline in the time needed to estimate of 7.24%, it is not sufficiently accurate when it comes to predicting extreme values or peaks over longer time frames.Time series sequence of the input water quality parameters prediction model is defined as x_0 . . . , x-Tx_T and the goal of the model prediction is to output prediction values $\hat{y}_0$, . . . , $\hat{y}_T$ for a period in the future based on the observed historical data. Our prediction model is to obtain the function F and make it produce the following map:

$$\hat{y}_0, \ldots, \hat{y}_T = \mathrm{F}(\mathrm{x}_0, \ldots, \mathrm{x}_T) \quad (7)$$

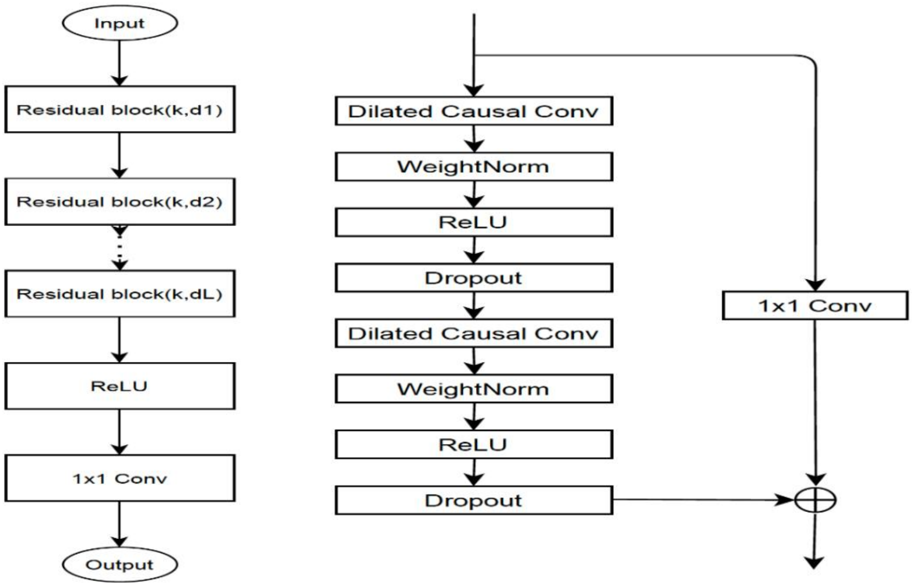

Figure 2. (a). Overall flowchart of TCN (b) TCN Residual Block (Fu et al., 2021)

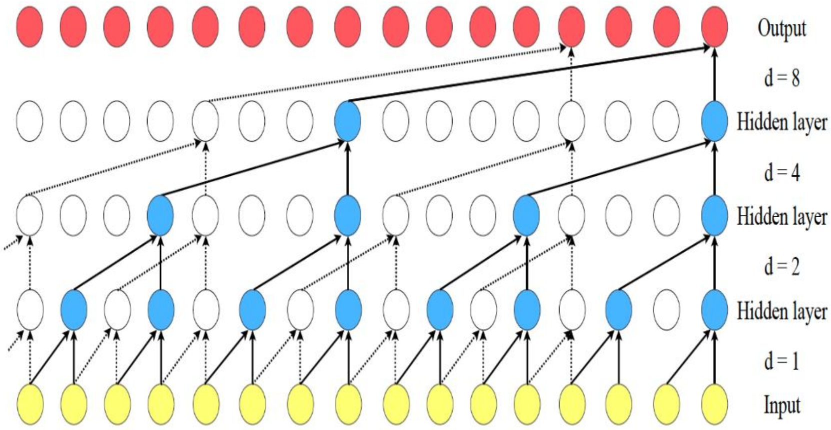

Figure 3. Dilated Causal Convolution (Fu et al., 2021)

The most common TCN residual block structure, which includes Dropout regularization, ReLU activation function, Weight Normalization, and dilated-causal convolution, is displayed in Figure 2 (b). If x is the residual block's input, then the residual block o's output can be defined like this:of the residual block, the output of the residual block o can be expressed as follows:

$$o = \text{Activation}(x + F(x)). \quad (8)$$

where Activation is the activation function and F(x) is the residual. The deep learning network's performance at learning will not decline as the stacked layers are constantly discovering new features, simply because the residual F(x) doesn't seem to be zero in practice. The Wave Nets network formed an early prototype for causal convolutions. Casual Convolutions were originally proposed in the Wave Nets network. Since the traditional CNN model cannot directly deal with the sequence problem, causal convolution can abstract the sequence according to $X_1, X_2, \dots \dots X_t$ and $y_1, y_2, \dots \dots y_{t-1}$ to predict y_t and make it close to the actual value. Causal convolution models, as compared to recurrent neural networks (RNNs), do not employ recurrent connections, allowing time series data to be input in parallel. TCN employs Dilated-Causal Convolution (DCC) to enhance the receptive field of neurons without significantly increasing computing cost, and it combines Dilated Convolution with causal Convolution to address the issue of standard causal Convolution's limited receptive field. The one-dimensional dilated causal convolution operation is expressed as follows:

$$F(S) = \sum_{i=0}^{K-1} f(i) x_{s-di} \quad (9)$$

where x is the input sequence, f (i) is the filter, also known as the convolution kernel, d is the dilation factor, k is the size of the convolution kernel, s − di ensures that only past inputs can be convolved.

2.4 Related work

When attempting to forecast essential water quality indicators, such as pH, water temperature, and dissolved oxygen. Liu et al. (2020) created an Accurate Prediction Scheme of Water Quality in Smart Agriculture by using Bi-S-SRU Learning Network. The Bi-S-SRU model was trained using the pre-processed data and correlation priors, demonstrating improved prediction accuracy compared to traditional methods. It showed improved prediction accuracy but had higher time complexity compared to RNN and LSTM methods, limiting real-time applications. Hu et al. (2023) suggested a hybrid deep learning model as an innovative approach for predicting the quality of river water. They enhanced the traditional Temporal Convolutional Network (TCN) with a self-attention mechanism and adjusted the residual block structure to accommodate the difficulties of water quality data, such as periodicity, seasonality, and nonlinearity. This model allows for the simultaneous updating of all layer weights at each time step, greatly increasing the computational efficiency of the model. TCNs utilize causal convolution, ensuring that predictions are based solely on past and present information, which is crucial for real-time forecasting in aquaculture. To capture the complex relationship between different water quality parameters, they used self attention mechanism. Since self-attention only computes a single weighted sum of all input data, so failed to capture complex, multidimensional correlations between water quality parameters and is computationally expensive due to the quadratic complexity of calculating attention scores.The comparison of various previous research, focusing on their methodologies and limitations, is presented in Table I.

Table I. Comparison of Previous Research

Paper	Research Objective	Methodology	Conclusion	Limitations
(Liu et al., 2020)	To predict water quality parameter i.e pH, water temperature and DO	Deep Bi-S-SRU Learning Network for time series data	Improved estimation accuracy for both short-term and long_term prediction	Higher in time complexity and only predicts for the next 3 days with an accuracy 94.42%.
(Fu et al., 2021)	To predict the water quality for the long term (i.e for the next 7 days) and reduce the cost of time for training and prediction	Temporal convolution network(TCN) with pearsons coefficient for feature selection.	Achieved accuracy(91.91%) for long-term prediction(7 days) and reduced time complexity for training and prediction by an average 64.92% and 7.24%, respectively.	This system is not accurate enough for long-term peak prediction, and failed to capture the complex relationships
(Zuo et al., 2023)	To address the issue of high temporal correlation and extreme fluctuation of	A combined model for water quality prediction based on VMD-TCN-ARIMA	Gained more precise and faster prediction results, improved the generalizability and intelligence of the model.	Poor performance for long-term prediction

	water quality attributes.	optimized by WSWOA		
(Hu et al., 2023)	To predict the water quality for a longer period of time	A hybrid improved temporal convolution network model with self-attention	Achieved higher prediction accuracy than other models, forecasted water quality parameters of the river over a longer period more accurately.	Limited to a single focus and did not capture diverse relationships.
(Palabıyık & Akkan, 2024)	To estimate the WQI index using a statistical approach and other AI models	Multiple Linear Regression, Multi-layer perceptron and other machine learning algorithms	The MLR model performed with high accuracy, with R^2 value of 1.0 and an RMSE value of 0.0025, compared to ANN and ML	Limited to linear relationships and difficult to determine the individual effect of a variable in the case of multicollinearity.

3. Methodology

3.1. Research Methodology

Figure 4 depicts the designed research methodology to predict the water quality using TCN with a multi-head attention mechanism.

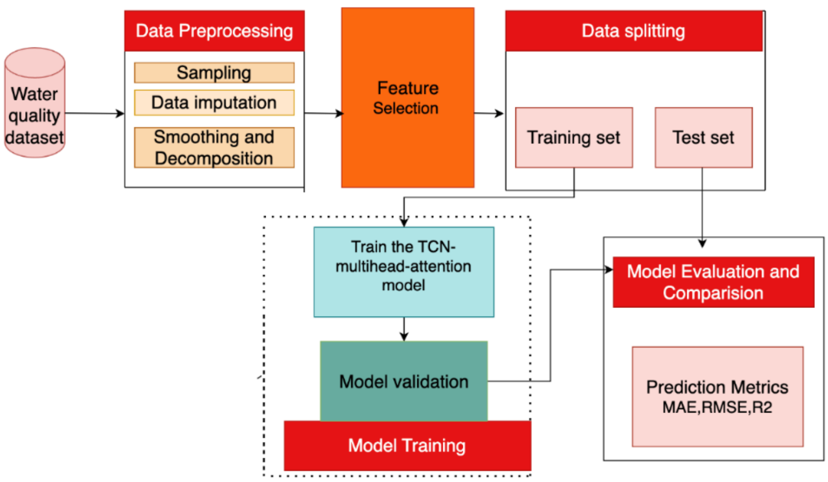

Figure 4. Proposed block diagram for water quality Prediction using TCN with multihead

3.2. Data Acquisition

The original dataset, which was captured at 30-minute intervals at the Burnett River monitoring station between 2013 and 2021, is available through the Queensland Government Open Data portal. The dataset was initially a multiple-column dataset without subgroups for testing and training. This dataset, which includes approximately 140000 samples overall, encompasses seven water quality parameters: temperature, PH, dissolved oxygen, electrical conductivity, turbidity, dissolved oxygen saturation, and chlorophyll content.

3.3. Data Pre-processing

The dissolved oxygen feature is isolated from the original dataset and used for down sampling to resample all of the data every day. After resampling, the data had missing information; therefore imputed the missing information was imputed using the forward fill approach. Certain characteristics of the time series have been extracted by using the SG filter algorithm in order eradicate the high frequency fluctuations, thus enhancing the prediction of the future water quality indicator value. Then, the smoothed data are divided into three sub-series: residual, trend, and seasonality by using the STL method. This enhanced the interpretability of the data by facilitating the discovery of underlying patterns, such as long-term trends and periodic effects.

3.4. Features Selection

Feature selection is the process of finding and eliminating redundant, insignificant, and unimportant features from raw data that don't help with the accuracy measure evaluation. It assists in reducing extra computing complexity, achieving the ideal amount of features, and improving the predictive model's accuracy. Relationships between features and the target variable were examined using Pearson's correlation. In order to choose features, features with an absolute correlation higher than the 0.1 criterion were chosen, while features with strong multicollinearity, defined by a threshold of 0.8, were eliminated. There were seven water quality measurements at first, but after decomposing, it became 23, and only 14 columns of parameters were selected by using Pearson's correlation.

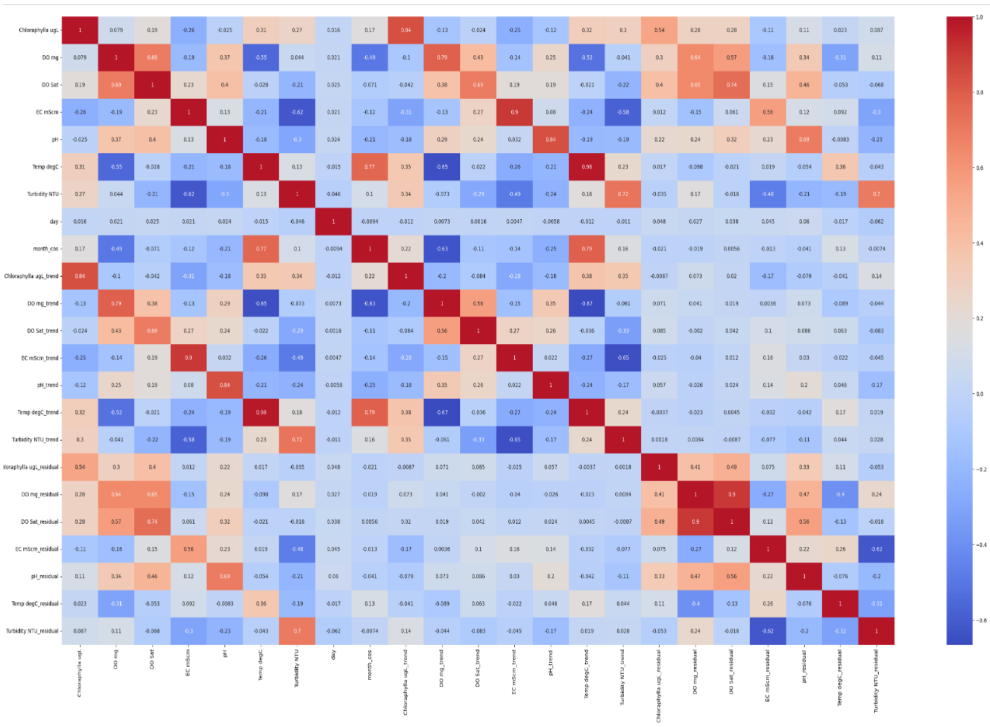

Figure 5. Correlation Matrix before feature selection

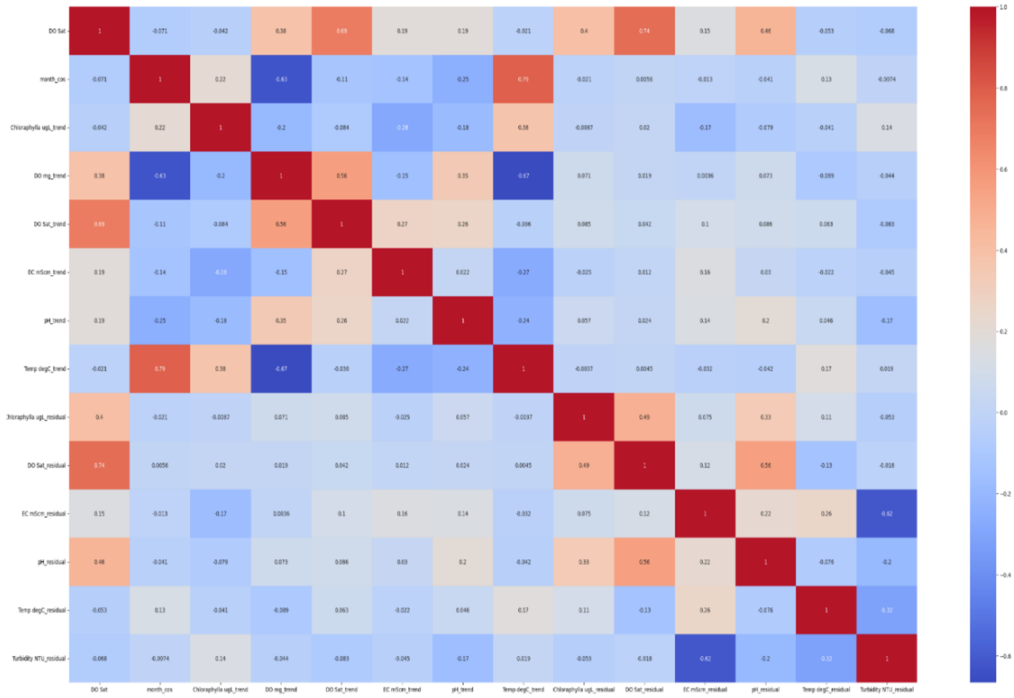

Figure 6. Correlation matrix of selected features

3.5. Model Development

3.5.1. TCN with Multihead attention

Once the data preparation is completed, the model development phase focuses on designing architectures tailored for water quality prediction. In this study, TCN is enhanced with a multihead-attention mechanism.

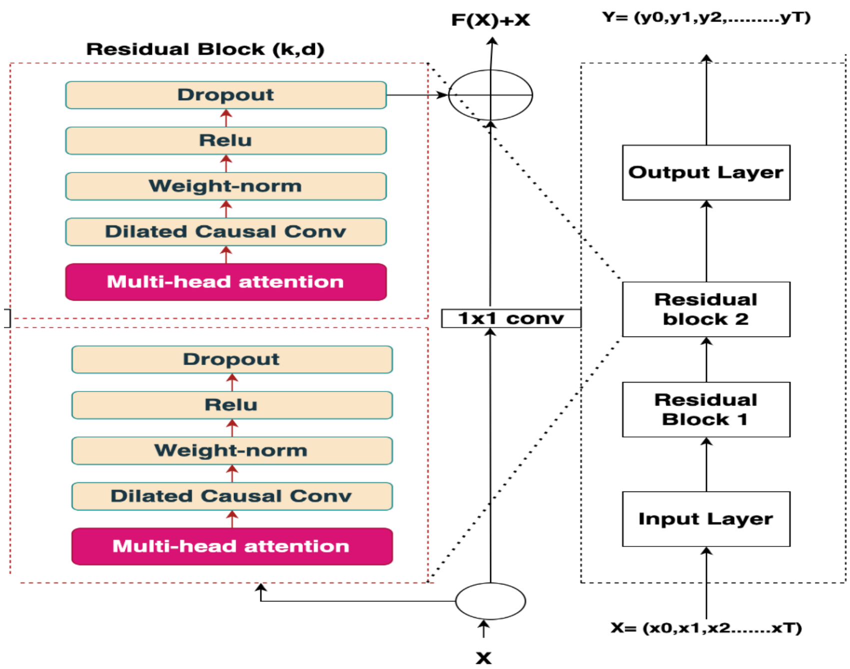

Figure 7. (a) Residual block structure of TCN with multihead attention, (b) TCN network structure (Tong et al., 2022).

To further refine the prediction capability, the TCN with a multihead attention mechanism introduces an attention layer, enabling the model to dynamically focus on the most significant temporal features in the dataset. This model is trained on the training dataset and evaluated against test data to predict water quality parameters, ensuring robust performance while addressing the challenges posed by complex temporal dependencies.

The Multi-head attention mechanism is an expansion of the attention model by introducing multiple heads, which aims to capture the internal correlation of the data, so as to further improve the prediction ability of the model (Tong et al., 2022). The designed structure of the multihead-attention mechanism is shown in Figure 8, which consists of three vectors: Query (Q), Key (K), and Value (V), which are generated from input data. These vectors are projected into multiple sets (heads) using linear layers with different weights.

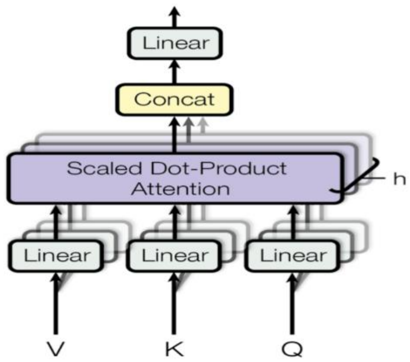

Figure 8. Multi-Head Attention mechanism (Tong et al., 2022)

Instead of linearly projecting queries, keys, and values into a single subspace, it projects them into multiple subspaces, computes similarity, and applies the attention function in parallel (Tong et al., 2022). As shown in Eq. (10), the resulting vectors are concatenated and mapped again to obtain the final output.

$$\mathrm{MultiHead}\ (Q, K, V) = \mathrm{Concat}\ (\mathrm{head}_1, \ldots.\mathrm{head}_2, \ldots..\mathrm{head}_h)W^0, \tag{10}$$

$$\text{where } \mathrm{head}_i = \mathrm{Attention}(QW_i^Q, KW_i^K, VW_i^V)$$

3.5.2. Model Evaluation and Validation

The dataset is divided in half, with 80% going to the training dataset and 20% going to the test dataset. Further, we divided the training data into a training set and a validation set. To effectively examine our proposed model, Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R- squared (R^2) are used.

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{11}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(|y_i - \hat{y}_i|)^2} \tag{12}$$

$$R^2 = 1 - \frac{\sum(y_i-\hat{y}_i)^2}{\sum(y_i-\bar{y})^2} \tag{13}$$

The method used for validation is Time-series cross validation based on an expanding window. Start with a small subset of data for training purposes, forecast for the later data points and then checking accuarcy for the forecasted data points. The same forecasted data points are then included as part of the next training dataset, and subsequent data points are forecasted.

4. Results and Discussions

4.1. Dissolved Oxygen Prediction for the next day using TCN

The TCN model was trained with training data and validated using the validation data. We first predict the dissolved oxygen for the next 24 hours using the TCN model with the optimized hyperparameter values. In our case, we employ the first 7 days of water quality parameters as model inputs, and the predicted value for the next day is the output, which means the sliding window size is 7 and the prediction step size is 1. Time-series cross-validation based on an expanding window is implemented to validate the model and make it more effective. Figure 9 illustrates the training and validation loss of a TCN model during its training process.

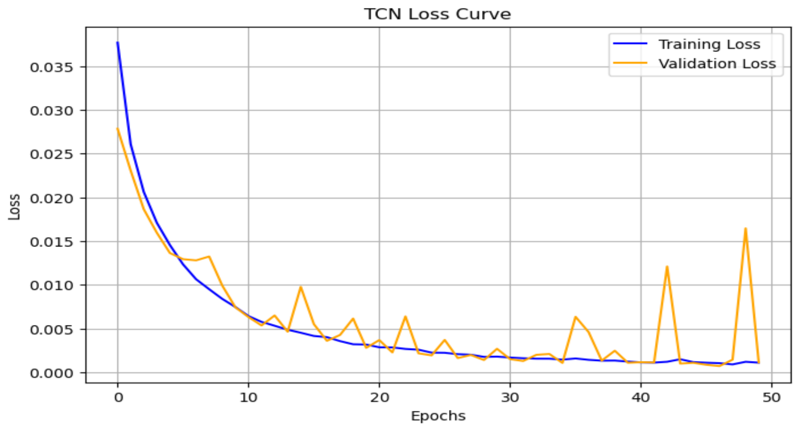

Figure 9. Validation loss vs Training loss using TCN for next-day prediction

As previously mentioned, the reliability of the predictive models is demonstrated by the reduced MAE and RMSE as well as greater R^2 throughout all data sets. The TCN model measured the value 0.1415 of RMSE and 0.1247 of MAE. By analyzing these performance evaluation values, we can conclude that this model performed acceptable results.

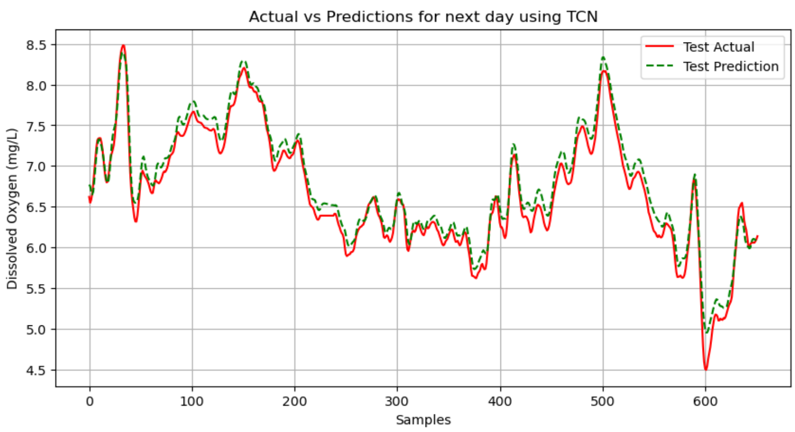

Figure 10. Prediction of DO for the next day using TCN

4.2. Dissolved Oxygen Prediction for the next day using TCN with Self-attention

Figure 11 shows a loss plot using a Temporal Convolutional Network (TCN) model with self-attention. After performing the validation technique, the training loss usually falls as the model passes through the training epochs, demonstrating that the model is getting better at matching the training data.

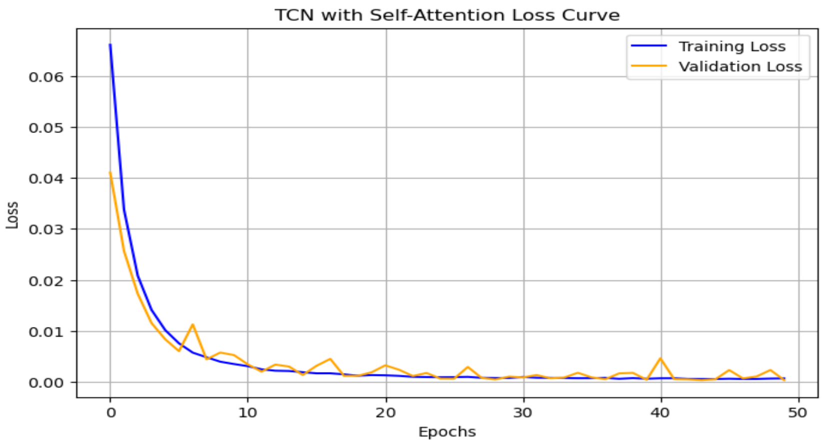

Figure 11. Validation loss vs Training loss using TCN with Self-attention for next day prediction

By using bayesian optimization technique, our proposed model is trained on 50 epochs and then predicts the dissolved oxygen for the next time step. This prediction evaluates the RMSE value of 0.1114, which is better than the TCN models. From Figure 12, it is clear that TCN with self-attention performed better than TCN and the LSTM model.

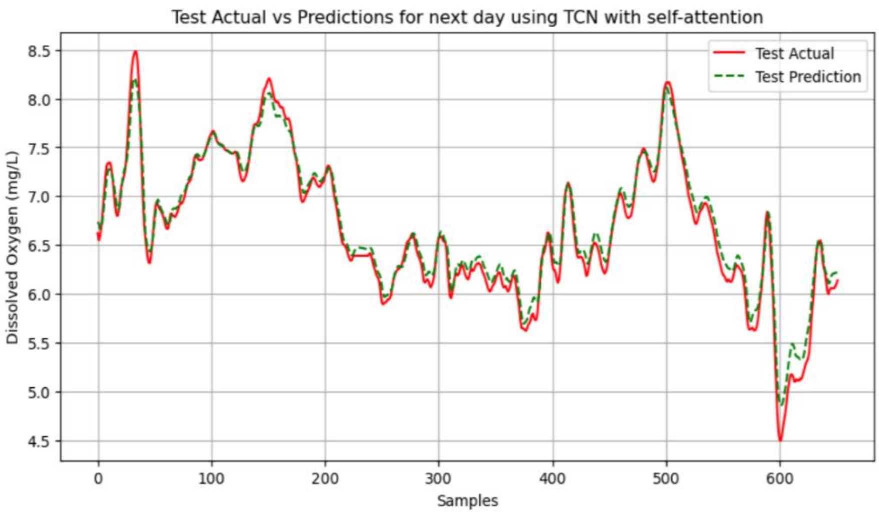

Figure 12. Prediction of DO for the next day using TCN with Self-Attention

4.3. Dissolved Oxygen Prediction for the next day using TCN with Multihead attention

Figure 13 shows, as the model advances through the training epochs, the training loss typically drops, suggesting that the model is improving its ability to match the training data. The validation loss demonstrates how well the model works with unknown data.

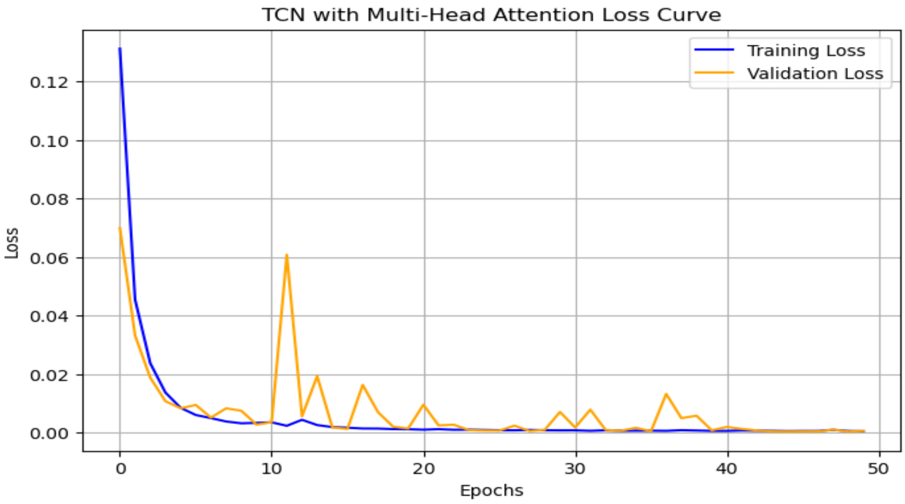

Figure 13. Validation loss vs Training loss using TCN with Multi-head Attention for next day prediction

The combined model TCN-Multihead-attention, which predicts the water quality elements 'DO' with acceptable fitting accuracy in the prediction and preserves good robustness in the context of rapidly fluctuating data, as shown in Figure 14. The evaluation metric MAE and RMSE of this prediction are 0.0457 and 0.0707, respectively. Also, the R^2 value is 99.06, which is closer to 1, that repsesents the better prediction accuracy. Overall, the results indicate good predictive performance, with room for improvement in capturing rapid changes through model optimization or advanced techniques.

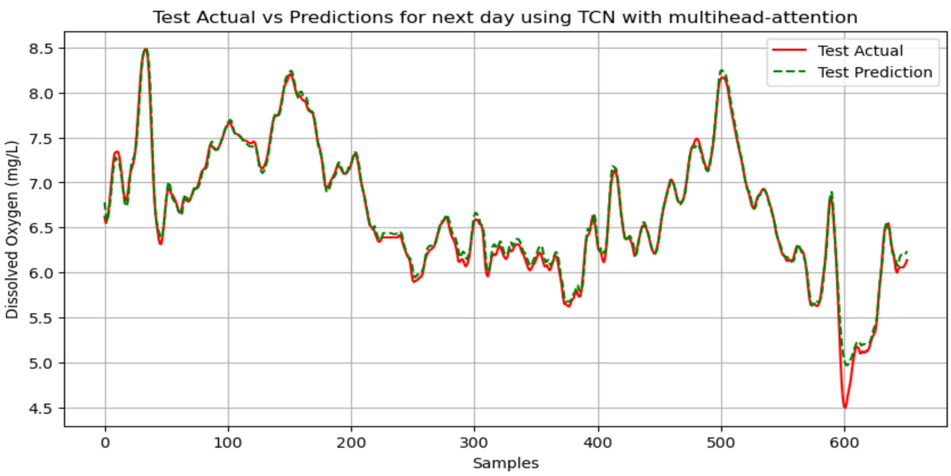

Figure 14. Prediction of DO for the next day using TCN with multi-head Attention

4.4. Dissolved Oxygen Prediction for the next day using LSTM

Figure 15 displays the training result with respect to validation loss after we ran the model with 50 epochs and a validation split.

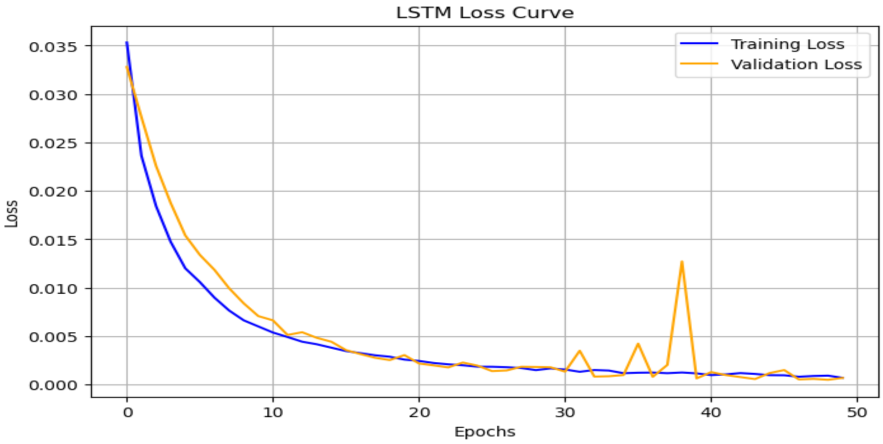

Figure 15. Validation loss vs Training loss using LSTM

Figure 16 illustrates the performance of a predictive model for dissolved oxygen (DO) levels for the next few days. The model did't captured the complex trends and rapid fluctuations in DO levels.

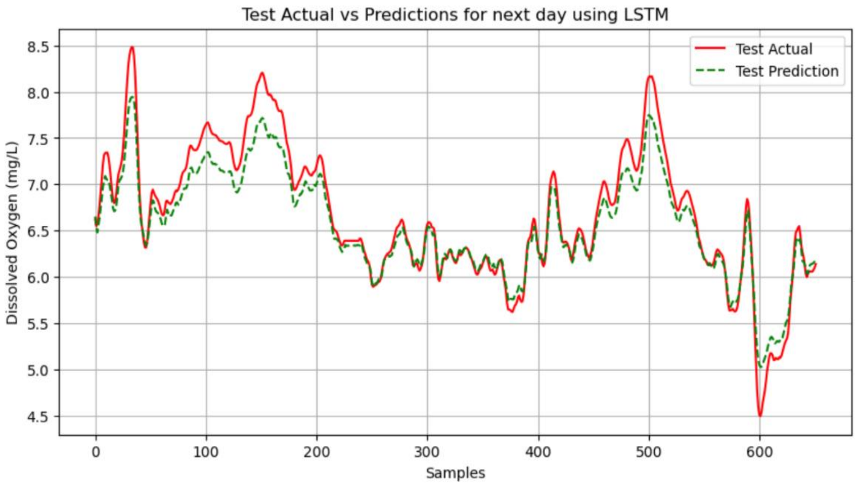

Figure 16. Prediction of DO for 1 days Using LSTM

4.5. Predicting DO using Statistical Method (Multiple Linear Regression) for the next day

To predict the water quality using a statistical method, Multiple Linear Regression(MLR) is used to predict the future values by using historical water quality parameters. The MLR technique is comparatively simple and takes less time because of its realistic return (Palabıyık & Akkan, 2024) .

The mathematical model of MLR is represented as:

$$y = \beta_0 + \beta_0 X_1 + \beta_0 X_2 + \ldots\ldots\ldots \beta_P X_P + \epsilon. \qquad (14)$$

where,

y = target variable

β_0= Intercept (the predicted value of y when allX_i=0

$\beta_{1,}\beta_2,\ldots\ldots\beta_p$ =coefficient associated with each independent variable

$X_1, X_2, \ldots\ldots X_p$ = feature variable

ϵ = Error value (it is the difference between the actual value and the predicted value of taget feature)

The plot demonstrates the statsitical approach and is shown in Figure 17. This predicted R^2 value of 0.9638, and RMSE value of 0.13, which proved that this model can pedicts future data with average performance. By comparing stastistical approach with the AI model, the proposed AI model was more desirable due to its performance. However, this statistical model is easy to implement and has reduced time complexity than an AI model.

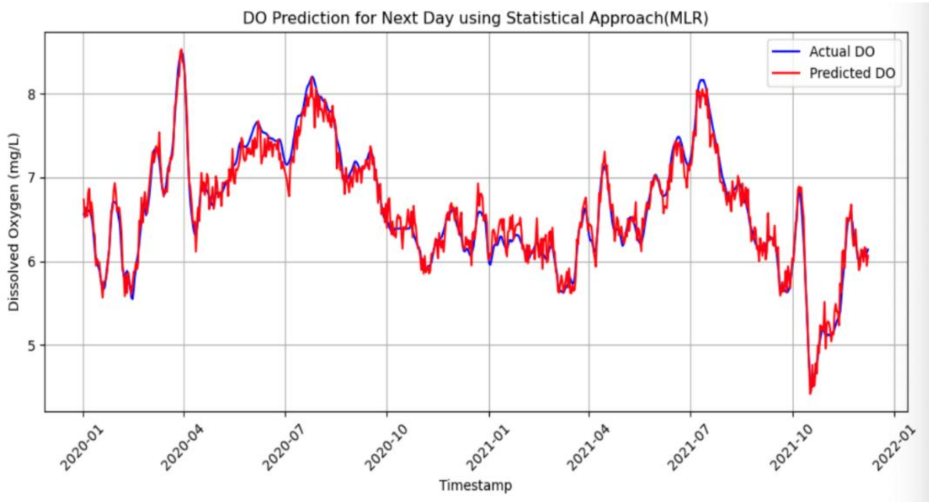

Figure 17. Prediction of DO for 1 day using MLR

4.6. Dissolved Oxygen Prediction for the next 7 Days using TCN

In this case, we used a sliding window of size 14 and predicted days equal to 7 days for forecasting. To validate this method expanding window cross-validation technique was applied and iterated over 50 epochs. The result of training and validation loss is displayed in Figure 18.

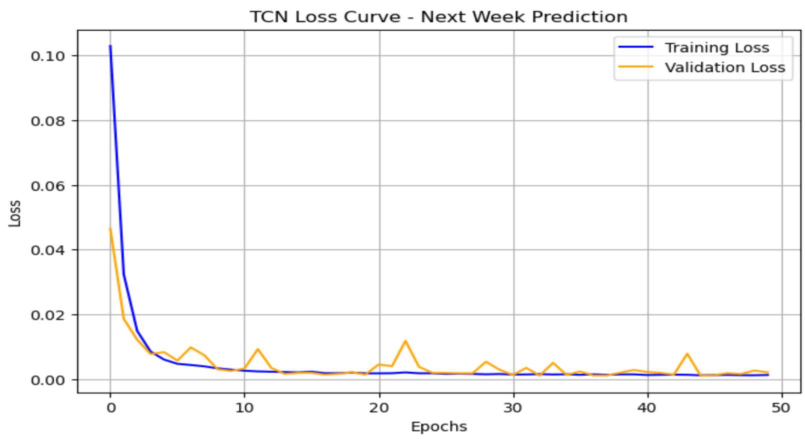

Figure 18. Validation loss vs Training loss using TCN for next week's prediction

The error between actual and prediction within the test dataset was measured by using techniques such as MAE and RMSE value, which are shown in Figure 19, and their values are 0.1592 and 0.2069, respectively. This shows that this model performed average performance for long-term prediction as it did not capture complex and highly fluctuating dissolved oxygen values.

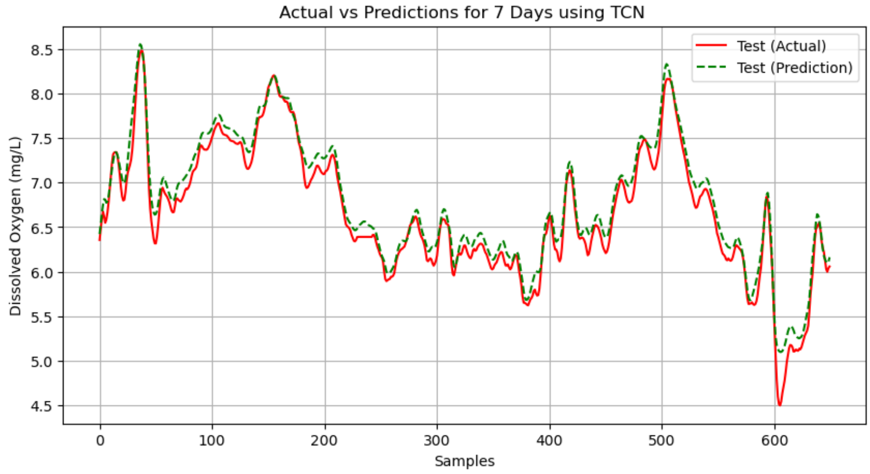

Figure 19. Prediction of DO for 7 days using the TCN model for 7-day prediction

4.7. Dissolved Oxygen Prediction for the next 7 Days using TCN with self-attention

Figure 20, it shows a gradual decrease in both Validation and training loss with increasing epoch and it performed reduced validation loss when compared to TCN model. This prediction model learns slowly and stops learning after 20 epochs.

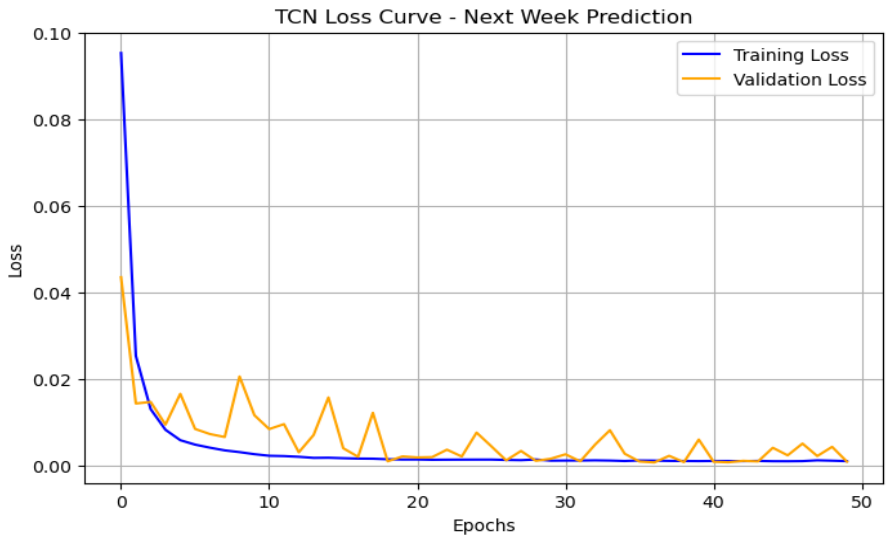

Figure 20. Validation loss vs Training loss using TCN and self-attention for next week prediction

To predict the dissolved oxygen for next week, 40 epochs are used for TCN with self attention model, and the total number of residual blocks used is 4, and the learning rate is 0.0001. The value of MAE and RMSE is 0.1458 and 0.1758, respectively. This model gained higher accuracy than the TCN model.

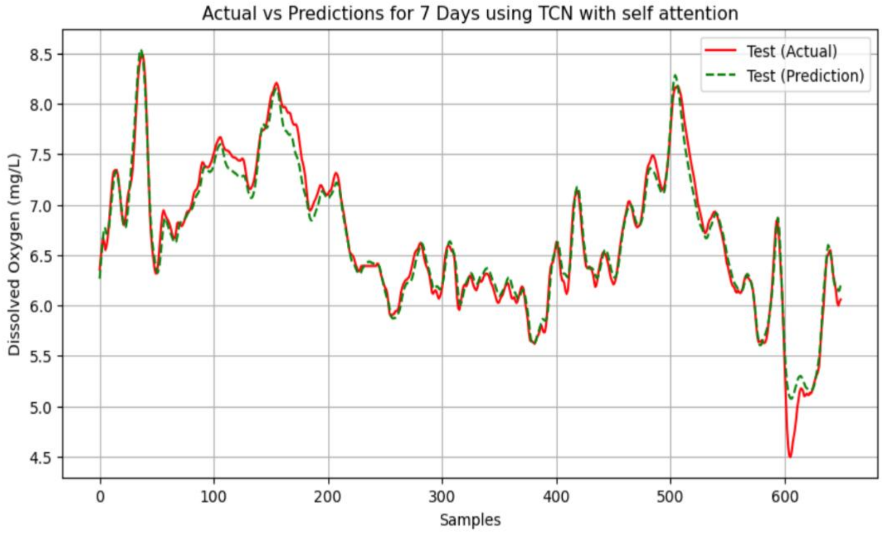

Figure 21. Prediction of DO for 7 days using TCN with self-attention

4.8. Dissolved Oxygen Prediction for the next 7 Days using TCN with multi-head attention

Firstly, three heads have been used to capture the complex structure of different water quality elements and later applied to the dilated causal convolution as an input. There was total of 4 residual blocks have been used in this model, and a batch size of 64 with L2 regularization of 0.00013.

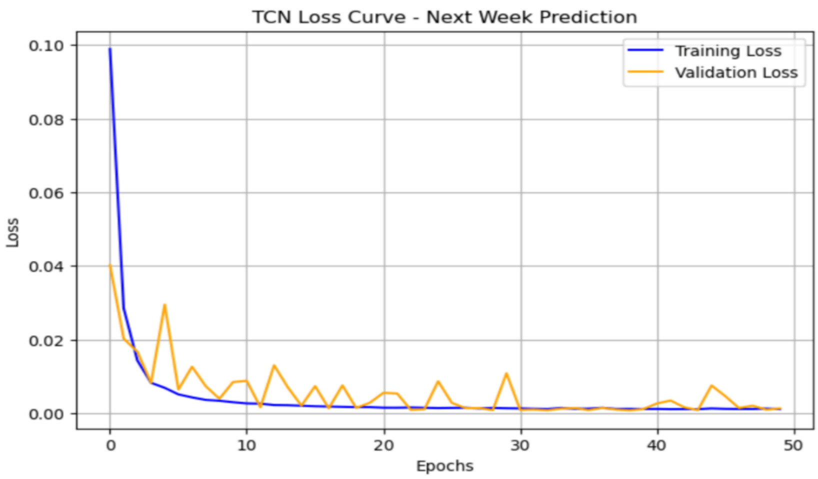

Figure 22.Training and validation loss using TCN with multi-head attention for the next 7 days prediction

The forecasting outcomes from the test dataset are presented in Figure 23 R^2 values of 0.9651 and 0.1005 of MSE, it shows that the results of TCN with multihead attention are closer to the real observations for the long term. It seems that compared to other models, this one is more accurate and remains better for long-term predictions.

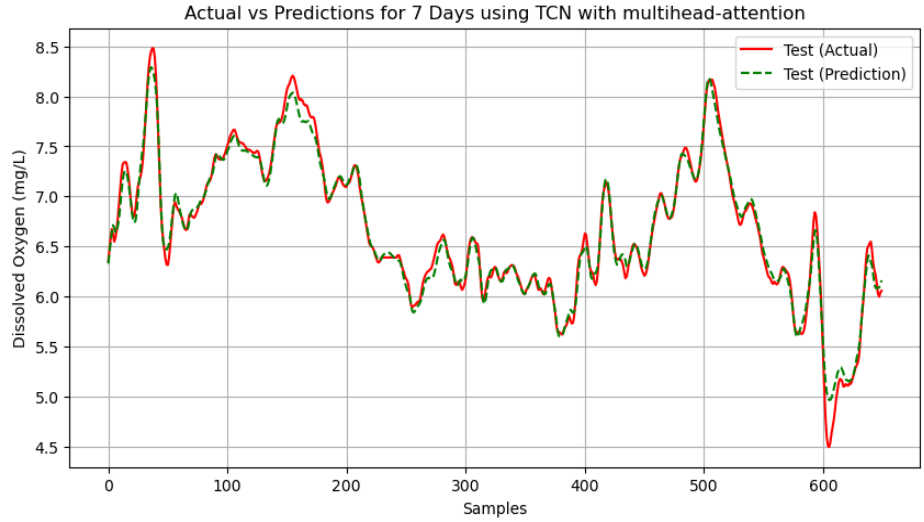

Figure 23. Prediction of DO for 7 days using TCN with multi-head attention

4.9. Dissolved Oxygen Prediction for the next 7 Days using LSTM

Figure 24 is the loss plot of training and validation model. Similar to next day prediction, it has also been trained using the same number of epochs and the same activation function, which is called the sigmoid function. This model obtained 0.0030 of training loss and 0.0019 of validation loss.

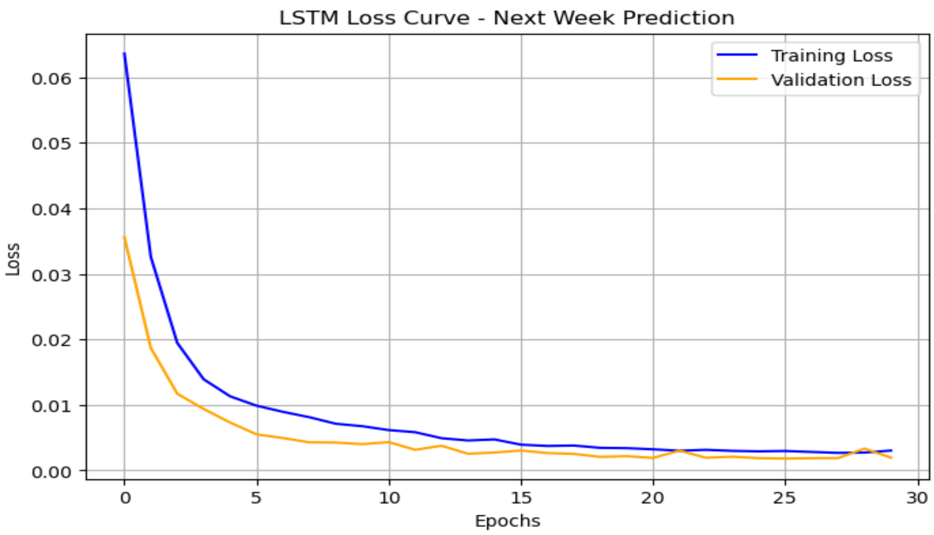

Figure 24. Training and validation loss using LSTM for the next 7 days' prediction

The deviation between actual and predicted water quality data is greater, with an RMSE value of 0.2433 and R^2 value of 0.8892. The LSTM model didn't predict accurately for the long term since it has problems of capturing long-term relationships between various water quality parameters.

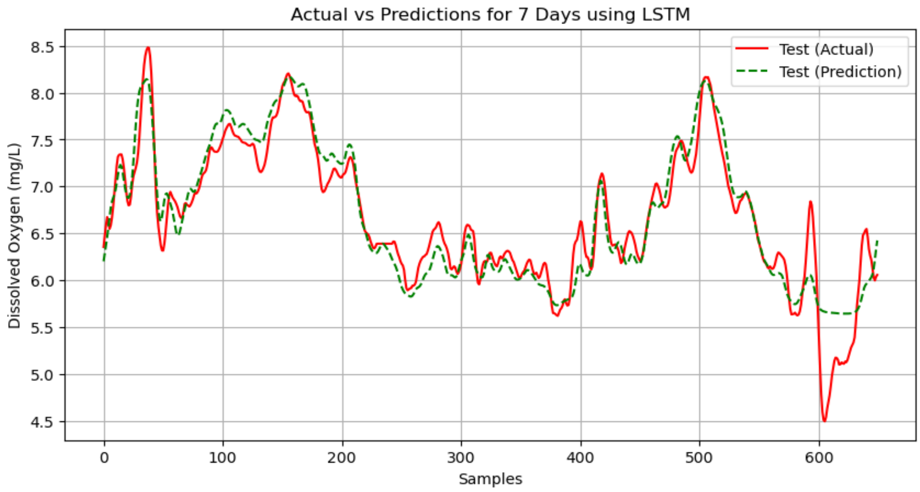

Figure 25. Prediction of DO for 7 days Using LSTM

4.10. Predicting DO using Multiple Linear Regression for the next 7 days

Forecasting of DO for the next seven days, MLR model measured performance metrics with MAE of 0.1970, RMSE of 0.2506 and R^2 of 0.888, which was low compared to other AI models.

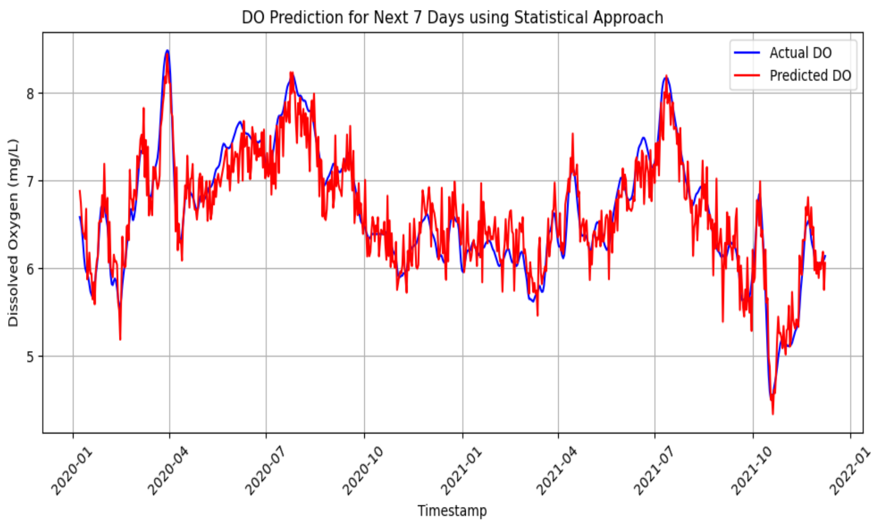

Figure 26. Prediction of DO for 7 days Using Multiple Linear Regression Model

4.11. Model Comparison and Discussion

As a result, for different AI models such as LSTM, TCN, TCN with Self-attention, TCN with multihead-attention and statical approach that are MLR mechanism are obtained from the training, the result is seen as very competitive. From the loss point of view, adding the multihead attention mechanism improved learning efficiency and reduced overfitting, resulting in faster convergence and better performance. The Figure 27 and Figure 28 illustrate the performance of five different models (MLR, LSTM, TCN, TCN with Self attention and TCN with Multi-head Attention) in terms of Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) across varying prediction horizons, respectively. The comparative analysis of

LSTM, TCN, TCN with self-attention and TCN with Multihead attention model for predicting the Dissolved Oxygen yielded insightful findings. The TCN model with multihead attention exhibits better performance for long-term prediction than several other commonly used benchmark models.

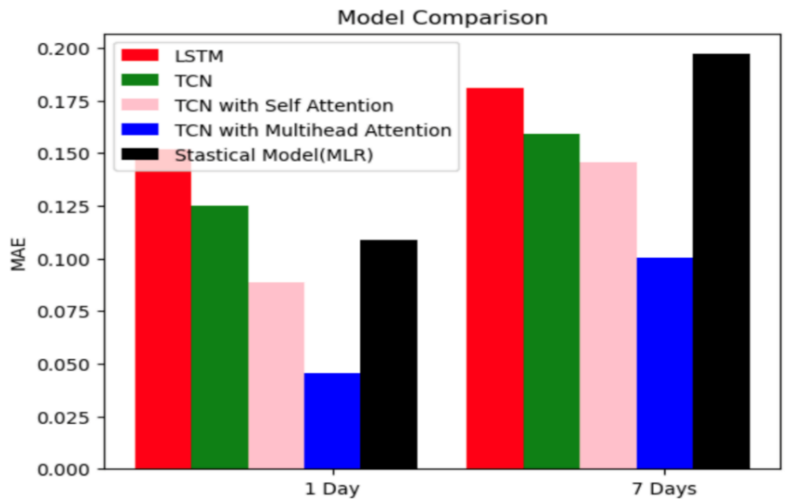

Figure 27. Total MAE of TCN with the Multihead attention model and other baseline models

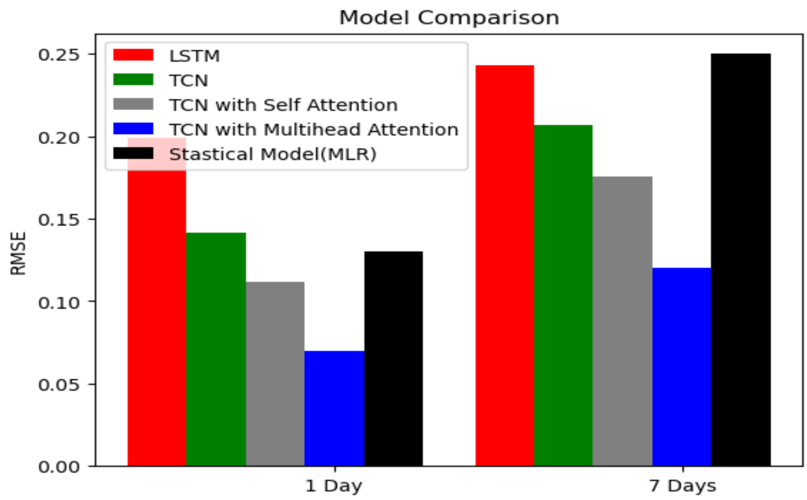

Figure 28. Total RMSE of TCN with the Multihead attention model and other baseline models.

Table II shows that **TCN with Multi-head Attention** outperformed other models in terms of predictive accuracy, especially for longer-term predictions. This suggested that the multi-head attention mechanism enhances the TCN's ability to capture complex patterns and dependencies in the data.

Table II. R^2 values of different models

$Performance\ Metrics: R^2$

Days	TCN	TCN with Self-attention	TCN with Multihead-attention(proposed)	LSTM	MLR
Day 1	0.9625	0.9768	**0.9906**	0.9260	0.9638
Day 7	0.9199	0.9421	**0.9651**	0.8892	0.8805

5. Conclusion and Future Recommendation

5.1. Conclusion

Accurate forecasting of water quality is crucial for protecting aquatic ecosystems, yet long-term forecasting remains challenging due to the nonlinear and dynamic characteristics of water quality parameters. In the study, dissolved oxygen is finally predicted for short-term and long-term (1 Day and 7 Day) using a hybrid TCN with Multihead Attention, consistently obtaining the maximum accuracy, followed by the combined TCN with self-attention and TCN, according to the results. MLR and LSTM show the highest RMSE and MAE values for next week, which indicates worse accuracy. Based on the experiments, the performance evaluation metrics show slight variations in accuracy when assessed on one day and seven days of prediction. Value of MAE for LSTM, TCN, TCN with self-attention, TCN with multihead attention and MLR model are 0.1516, 0.124, 0.0888, 0.0457 and 0.1089, respectively, for next day prediction and 0.1808, 0.1592, 0.1458,0.1005 and 0.1970 for next week for the same case. In contrast of R^2, value of R^2 for LSTM, TCN, TCN with self-attention, TCN with multi-head attention and MLR model are 0.9260,0.9625,0.9768,0.9906 and 0.9638 for next day and 0.8892,0.9199,0.9421,0.9651 and 0.8805 for next week. The results demonstrated that by integrating Multi-head Attention with TCN outperforms better performance than other models, achieving the lowest prediction errors and the highest R-squared values for prediction horizons (day 1 and day 7). From the above experiment, we also concluded that the statistical method can predict water quality values for the short term with good accuracy, whereas for long-term prediction, this approach is not suitable. So, an AI model is best for long-term prediction with high efficiency and generalization. The performance of all models tends to degrade as the prediction horizon increases from day one to day seven, which is expected due to the inherent uncertainty in long-term forecasting. However, our proposed model predicts water quality data for longer-term prediction with higher accuracy than other models. These findings conclude that incorporating a multihead attention mechanism with TCN can perform superior performance across various prediction horizons, so that administration can find water-related illnesses early and take corrective measures before they occur to address the issue and better safeguard the aquatic environment.

5.2. Limitations and Future Recommendations

The future of this research has a lot of scope for development. One of the research's limitations involves the fact that this approach only took historical water quality data as input, ignoring the influence of other significant factors. Secondly, this research has been spent on increasing model performance rather than decreasing model time complexity, which has a significant memory and computational cost, which rises quadratically with the length of the sequence. In future, we can consider the impact of different external conditions, such as rainfall and climate, as inputs, and the model's usability in real-world settings can be further improved. A variety of methodologies can be applied, including hyperparameter tuning and assessing the model's applicability for other lakes, rivers, sea surfaces, etc.

REFERENCES

Bi, J., Lin, Y., Dong, Q., Yuan, H., & Zhou, M. (2021). Large-scale water quality prediction with integrated deep neural network. *Information Sciences*, *571*, 191–205. https://doi.org/10.1016/j.ins.2021.04.057

Fu, Y., Hu, Z., Zhao, Y., & Huang, M. (2021). A Long-Term Water Quality Prediction Method Based on the Temporal Convolutional Network in Smart Mariculture. *Water*, *13*(20), 2907. https://doi.org/10.3390/w13202907

Haghiabi, A. H., Nasrolahi, A. H., & Parsaie, A. (2018). Water quality prediction using machine learning methods. *Water Quality Research Journal*, *53*(1), 3–13. https://doi.org/10.2166/wqrj.2018.025

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*(1), 11260. https://doi.org/10.1038/s41598-023-38465-3

Li, X., Sha, J., & Wang, Z.-L. (2017). Chlorophyll-A Prediction of Lakes with Different Water Quality Patterns in China Based on Hybrid Neural Networks. *Water*, *9*(7), 524. https://doi.org/10.3390/w9070524

Liang, K., & Zhang, C. (2024). *Integrated Approaches in Water Quality Prediction Via Tcn-Lstm-Attention Model And Reinforcement Learning*. https://doi.org/10.2139/ssrn.5050262

Liu, J., Yu, C., Hu, Z., Zhao, Y., Bai, Y., Xie, M., & Luo, J. (2020). Accurate Prediction Scheme of Water Quality in Smart Mariculture With Deep Bi-S-SRU Learning Network. *IEEE Access*, *8*, 24784–24798. https://doi.org/10.1109/ACCESS.2020.2971253

Morandini, S., Fraboni, F., De Angelis, M., Puzzo, G., Giusino, D., & Pietrantoni, L. (2023). The Impact of Artificial Intelligence on Workers' Skills: Upskilling and Reskilling in Organisations. *Informing Science: The International Journal of an Emerging Transdiscipline*, *26*, 039–068. https://doi.org/10.28945/5078

Palabıyık, S., & Akkan, T. (2024). Evaluation of water quality based on artificial intelligence: Performance of multilayer perceptron neural networks and multiple linear regression versus water quality indexes. *Environment, Development and Sustainability*. https://doi.org/10.1007/s10668-024-05075-6

Shams, M. Y., Elshewey, A. M., El-kenawy, E.-S. M., Ibrahim, A., Talaat, F. M., & Tarek, Z. (2023). Water quality prediction using machine learning models based on grid search method. *Multimedia Tools and Applications*, *83*(12), 35307–35334. https://doi.org/10.1007/s11042-023-16737-4

Tong, C., Zhang, L., Li, H., & Ding, Y. (2022). Temporal inception convolutional network based on multi-head attention for ultra-short-term load forecasting. *IET Generation, Transmission & Distribution*, *16*(8), 1680–1696. https://doi.org/10.1049/gtd2.12394

Zhou, J., Wang, Y., Xiao, F., Wang, Y., & Sun, L. (2018). Water Quality Prediction Method Based on IGRA and LSTM. *Water*, *10*(9), 1148. https://doi.org/10.3390/w10091148

Zuo, H., Gou, X., Wang, X., & Zhang, M. (2023). A Combined Model for Water Quality Prediction Based on VMD-TCN-ARIMA Optimized by WSWOA. *Water*, *15*(24), 4227. https://doi.org/10.3390/w15244227