International Academic Institute for Science and Technology

International Academic Journal of Science and Engineering
Vol. 13, No. 3, 2026, pp. 417-431.

ISSN 2454-3896

International Academic Journal of Science and Engineering

www.iaiest.com

Explainable Water Quality Index Forecasting Framework Using Hybrid Graph Attention and Transformer Networks for Intelligent Early Warning Systems

R. Sarala [a*], R.S. Ponmagal [b]

[a*] *SRM Institute of Science and Technology, Chennai, Tamil Nadu, India.*
E-mail: sr8749@srmist.edu.in, Orcid: https://orcid.org/0009-0007-9776-2413
[b] *SRM Institute of Science and Technology, Chennai, Tamil Nadu, India.*
E-mail: ponmagas@srmist.edu.in, Orcid: https://orcid.org/0000-0002-7228-1310

Abstract

The importance of water quality monitoring for aquatic ecosystem protection cannot be overemphasized, but the traditional machine learning algorithms used for monitoring data analysis have shortcomings in failing to account for the spatiotemporal interactions of the data. This study seeks to develop a spatiotemporal deep learning framework that employs Graph Attention Networks (GAT), Temporal Fusion Transformers (TFT), and SHAP-based interpretability for WQI forecasting and early warning purposes. Firstly, preprocessing of multisite physicochemical data involves filling in missing values using interpolation, removal of anomalies, and Min-Max normalization. Secondly, a spatial graph of monitoring stations is created, whereby the GAT layer captures the spatial information by allocating attention scores to the neighboring nodes, and TFT is employed to capture long-term temporal interactions and nonlinear trends. Lastly, a persistence-based early warning system identifies pollution degradation risks, while SHAP gives insight into parameter contribution. Empirical tests have revealed that the GAT-TFT model significantly outperforms conventional techniques (ANN, LSTM, Hybrid LSTM, and basic GNNs), exhibiting superior prediction efficiency with MAE = 0.048, RMSE = 0.079, and R^2 = 0.96. Moreover, the early warning system exhibits an accuracy of 0.92, precision of 0.89, recall of 0.94, an F1 score of 0.91, and a False Alarm Rate of 0.08, where dissolved oxygen, turbidity, and pH were found to be the most significant factors. In conclusion, the presented framework offers environmental agencies a powerful and highly efficient method for proactive management of water resources.

Keywords: Water Quality Index (WQI), Water Quality Monitoring, Spatiotemporal Prediction, Graph Attention Network (GAT), Temporal Fusion Transformer (TFT), Explainable Artificial Intelligence (XAI), SHAP.

DOI: 10.71086/IAJSE/V13I3/IAJSE13125
Received: 20 June, 2026/Revised: 23 July 2026/Accepted: 27 August 2026/Published: 15 September 2026

Introduction

Water resources play a crucial role in ensuring ecological balance, agriculture, and safe drinking water. Nonetheless, the quality of water in most areas has been greatly compromised by rapid industrialization, urbanization, and anthropogenic processes. Thus, effective environmental management and the use of water resources require continuous control and correct forecasting. The WQI is often employed as a powerful indicator of aquatic health, which incorporates a variety of physicochemical parameters--including pH balance, dissolved oxygen, cloudiness, biochemical oxygen demand, and chemical oxygen demand--into one measure of water condition, called the Water Quality Index (WQI). WQI prediction will allow the detection of the risk of contamination in time and preventive measures that can be implemented before it is too late to save the environment.

The previous research was mainly based on statistical analysis and standard machine learning. Machine learning models have emerged as one of the main standards for mapping the non-linear relationships between aquatic health variables and contamination indicators. As has been said, neural networks have been used to assess pollution in freshwater ecosystems, but hybrid models that use neural networks and optimization algorithms have been used to optimize dissolved oxygen predictions (Zulkifli et al., 2026; Yang, 2023). Despite the fact that these methods enhanced precision, the methods tended to treat monitoring observations as independent samples, and they were poor at taking into account the spatial relationships between monitoring stations. As deep learning has progressed, recurrent architectures and long short-term memory (LSTM) networks have been widely used to predict environmental temporal sequences. Created a better recurrent framework, which involves time-series prediction and spatial interpolation, and created an LSTM-based model with normalization strategies to increase the accuracy of the prediction (Wang et al., 2025; Mahesh et al., 2024). Developed an improved LSTM architecture, which combined empirical mode decomposition with LSTM to investigate non-linear and non-stationary properties of aquatic data. Regardless of these developments, the majority of architectures concentrate more on time relationships, and little attention is given to spatial relationships across observation sites. In response to this, hybrid deep learning models are developed (Lin et al., 2025). A hybrid LSTM-based model was introduced with optimization algorithms, and the reliability of predictions was enhanced by combining LSTM, XGBoost, and clustering (Chukwuemeka et al., 2026). Attention-based hybrid schemes combining fuzzy neural networks with bidirectional LSTM were proposed (Liu & Cheng, 2025). Although these models are able to increase the accuracy of predictions, highly complex spatial relationships between distributed monitoring stations are not adequately represented.

Nowadays, they have developed graph convolutional architectures as an effective tool for characterizing geographic interconnections. These models have the monitoring stations as nodes, and the edges are defined by spatial relationships such as geographic proximity or hydrological connectedness. A graph-based framework that combines convolutional and recurrent elements was proposed, adding attention modules to enhance dissolved oxygen predictions. Also proposed reinforcement learning-based graph neural networks to dynamically acquire spatial relationships (Wan et al., 2025; Ding et al., 2025). Also, joint capturing of spatial and temporal dynamics has become popular in spatiotemporal models. A chronological structure that could be adjusted based on spans was suggested, and a spatiotemporal graph convolution network that employs attention was proposed (Maqbool et al., 2026). Distant-horizon forecasting has also been demonstrated in the form of transformer-derived models because they have the potential to detect long-range correlations with the help of attention modules (Cui et al., 2025; Srinath & Tuppad, 2025; Zheng et al., 2024). Model interpretability is another important research direction. As deep learning models continue to increase in complexity, it is important to know how the environmental variables affect predictions. Transparency has been promoted using explainable artificial intelligence techniques. SHAP and LIME were used to analyze features and determine how environmental variables affected WQI predictions using robust models (Han et al., 2025; Nallakaruppan et al., 2024).

Moreover, IoT applications that monitor water quality and multimodal data have enhanced water quality prediction (Lv et al., 2025). Some studies suggested an IoT-enabled system to combine sensor networks with machine learning to perform real-time analysis, whereas others designed multimodal and spatiotemporal fusion models to predict in the long term. Although these have been made, a number of challenges still exist. Most of the models are concerned with only temporal prediction and have ignored the spatial correlation. Also, prediction accuracy is emphasized in most studies without considering early warning mechanisms, and there is a lack of model interpretability.

In order to overcome these constraints, this study suggests a composite spatiotemporal deep learning model to predict and give proactive warnings regarding the WQI. Graph Attention Networks (GAT) are used to capture geographical interrelations, and Temporal Fusion Transformer (TFT) describes chronological patterns. The hybrid model generates spatiotemporal representations that are complete in order to predict WQI. It has an early warning mechanism to catch possible water quality degradation and SHAP-based explainability to analyze the contribution of single parameters.

The principal contributions of the study follows (1), The hybrid GAT-TFT framework is created to consider both spatial and time-related dependencies, (2) the early warning mechanism is presented to detect the situation of possible water quality degradation, (3) the SHAP-based explainability module is included to enhance the transparency of the model, and (4) the proposed system offers an explainable and intelligent approach to water quality monitoring.

The paper is organized into six primary sections: Section I introduces the research background, limitations of existing models, and core objectives. Section II reviews related literature on water quality forecasting and explainable AI. Section III details the proposed GAT-TFT spatiotemporal methodology, early warning system, and SHAP framework. Section IV outlines the experimental setup, dataset statistics, and baseline models. Section V presents empirical results, model comparisons, and early warning performance. Finally, Section VI concludes the study.

Literature Review

The need to control water quality through proper prediction has become a significant field of study owing to the rising level of pollution in the environment and the need for sustainable management of water resources. Conventional methods of aquatic health appraisal generally use statistical methods and the implementation of standard automated learning methods to make predictions of ecological conditions using past data. ANNs have broadly been used to simulate nonlinear correlations amongst numerous physicochemical variables and water quality indicators. An example is that researchers applied artificial neural networks to determine parameters like dissolved oxygen, biochemical oxygen demand, and turbidity to measure the level of pollution in freshwater ecosystems (Zulkifli et al., 2026; Azma et al., 2023). Similarly, designed integrated automated learning systems, which combine neural structures with refinement processes that include biogeography-based metaheuristics and atomic pursuit algorithms to model daily dissolved oxygen concentrations. These methods showed better predictive accuracy than the conventional statistical techniques. Nonetheless, other traditional automated learning models assume that aquatic health measurements are independent observations and do not reflect the geographic interrelationships among sensing locations. As the concept of deep learning has matured, recurrent designs and long short-term memory (LSTM) have become popular in ecological temporal sequence prediction. These models are capable of learning intricate time-series patterns in sequential data and have demonstrated encouraging outcomes in water quality forecasting. The subsequent recurrent neural network framework was developed to better analyze the spatiotemporal changes in raw water quality through an integration of time-series prediction and spatial interpolation methods (Wang et al., 2025). Likewise, an LSTM-powered water quality prediction model was presented using a combined normalization approach to enhance prediction accuracy to aid in environmental monitoring (Perumal et al., 2023). An advanced LSTM model was proposed that uses refinement algorithms to refine parameter calibration and architectural convergence. Moreover, combined empirical mode decomposition and LSTM systems to reflect the non-linear and non-stationary nature of temporal aquatic data. Though these high-level learning methods have shown a remarkable enhancement in predictive power, the majority of them solely concentrate on time-dependent relations while neglecting the geographical interdependencies of observation locations (Lin et al., 2025).

In order to counter the constraints of the individual high-level learning architectures, combined schemes that combine various automated methodologies have been explored. Composite deep learning systems take advantage of the synergies between different operations to enhance predictive efficacy. Proposed a multifaceted architecture that combines LSTM models with heuristic search strategies (that is, gray wolf and fish swarm optimization) to predict aquatic health. Similarly, have designed a collective intelligence system that combines LSTM, XGBoost, and K-means grouping to obtain more accurate estimates and clear classification of environmental communication in arid areas (Chukwuemeka et al., 2026). Developed a multidimensional framework that integrates fuzzy logic systems and bidirectional LSTM and attention to optimize the operational capabilities of aquatic monitoring and proactive warning by fusing multi-sensor data. Although these have been improved, most hybrid models continue to have difficulties in capturing complex spatial relationships between monitoring stations in distributed monitoring systems.

The latest developments in graph neural networks have offered working solutions in modeling spatial dependencies in environmental monitoring systems (Zha & Yang, 2024). In graph-based models, the monitoring stations are depicted as nodes, whereas geographic proximity or hydrological connection between the nodes is depicted through the edges. This description allows the architecture to recognize the spatial correlations of ecological data, as well as dependencies between the positions of sensing. Designed a graph-based architecture and synthesized graph convolutional systems with gated recurrent units to extract both geographic and chronological information to predict multi-site aquatic health in fluvial systems. Improved the graph neural network structures by incorporating an indicator attention mechanism to predict dissolved oxygen better by determining the impactful environmental indicators. On the same note, suggested a reinforcement learning-based graph neural network model that learns dynamic adjacency relationships among monitoring stations to improve the accuracy of prediction. This research indicates that graph neural networks can be used to model the spatial dependencies of water quality monitoring systems.

Besides graph-based models, spatiotemporal deep learning models are also of growing popularity due to the fact that they can both represent spatial interactions and temporal dynamics in environmental data. Suggested a variable-length sequential prediction system that combines spatial embedding and temporal sequence learning to predict the water quality in several monitoring stations. Created an attention-based spatiotemporal graph convolution network, which considers spatial correlations and temporal dependencies via graph convolution and attention. Moreover, presented a transformer-based spatiotemporal graph integration system that employed both geographic and chronological attention modules to enhance the precision of distant-horizon aquatic forecasting. High-level learning architectures that are based on transformers have also shown considerable potential in the field of ecological projection, where they analyze sequences of time with large amounts of dependencies using attention-based systems. Presented a predictive model of water quality using transformers, combining fractional Brownian motion attributes with bottleneck transformation methods. Created a multi-stage transformer model for intelligent water quality monitoring. Also, introduced a joint high-level learning framework based on seasonal partitioning and attention units to optimize the accuracy of predictions of important aquatic health indicators. A second key research area is to enhance predictive model interpretability.

As deep learning models continue to get more complex, it is essential to understand how environmental parameters affect the prediction results, as this is vital to the process of environmental decision-making. Suggested an explainable artificial intelligence model based on SHAP and LIME methods to assess the quality of the water parameters in predictive outcomes. In line with this, employed powerful automated learning systems to forecast the Water Quality Index and determine how the ecological parameters affected the effectiveness of the forecast (Ding et al., 2023). These methods demonstrate the need to incorporate explainability methods to enhance transparency and reliability of the environmental monitoring systems. IoT-based monitoring systems and multimodal data fusion methods to enhance real-time monitoring of the environment have also been investigated by the researchers. Presented an IoT-based framework that incorporates the concept of sensing arrays and automated learning processes in order to measure and predict aquatic health variables in real time. Introduced a multi-source high-level learning structure, which merges temporal aquatic readings with atmospheric information via cross-modal integration techniques. On the same note, suggested a spatiotemporal multimodal fusion model that combines heterogeneous environmental data in predicting long-term water quality. These research papers show that sensor networks combined with environmental data and high-quality deep learning models can contribute greatly to the water quality monitoring system.

Despite these achievements, there are a number of challenges that still exist in current aquatic health forecasting processes. Most models concentrate on time prediction and ignore spatial correlation among the monitoring stations, as these are crucial in explaining the dynamics of water quality in distributed monitoring networks. Moreover, the majority of studies focus on the accuracy of prediction but fail to introduce early warning systems that can be used to actively identify possible pollution threats. Moreover, the complex deep learning models do not have high interpretability, and the environmental authorities struggle to explain what factors affect the results of the predictions. Thus, a comprehensive model, which captures spatial interactions, time-based dependencies, predictive forecasting, early warning, and explainable artificial intelligence, is needed to assist intelligent water quality monitoring systems.

Proposed Methodology

The current study suggests a comprehensible spatiotemporal high-level learning model to predict the Water Quality Index (WQI) and produce proactive notices in aquatic monitoring systems. This combination of Graph Attention Networks (GAT), Temporal Fusion Transformers (TFT), and SHAP-generated interpretability is effective in detecting geographical dependencies among sensing locations and time trends within ecological data. In contrast to traditional machine learning methods, where monitoring observations are considered to be independent samples, the proposed framework captures the interactions between monitoring stations in space in terms of a graph representation. Meanwhile, data on water quality is processed to learn patterns over time with the help of a transformer-based architecture.

The system has a number of stages in the overall workflow. To process water quality data, first, the missing values are treated, and features are normalized. Then, a spatial graph is created to reflect monitoring stations and relationships. Spatial dependencies are then extracted using Graph Attention Networks. These spatial dependencies are then transferred to a Temporal Fusion Transformer to learn time-series regularity. Lastly, the values of WQI are forecasted and assessed by an early warning system. SHAP explainability is used to explain the impact of environmental parameters on model predictions. Fig. 1 represents the general workflow of the proposed framework.

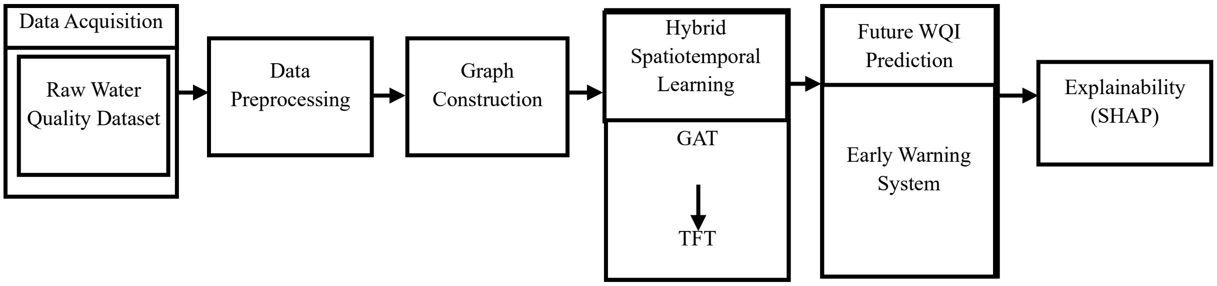

Fig. 1: Spatiotemporal Water Quality Index Prediction, Early Warning Detection, and SHAP

Data Preprocessing

The water quality dataset used in the current study is provided by the Water Quality Portal (WQP) on Data.gov, which combines the monitoring data offered by various environmental agencies. The dataset includes physicochemical parameters that are measured in various monitoring stations at different times. However, it is typical that environmental monitoring datasets may have missing values, noise, and inconsistencies because of sensor failure or unusually spaced sampling times. Thus, it is followed by preprocessing, followed by training of the prediction model. At this step, interpolation is used to fill in missing values, and statistical filtering is used to eliminate abnormal values shown in equations (1) and (2).

The dataset is normalized by the use of min-min max normalization so that all features can be of similar scale. Such a change can be defined as follows:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{1}$$

x is the original feature value, x_{min} and x_{max} are the minimum and maximum values of the feature, and x ' is the normalized value.

The Water Quality Index is not directly measured in the dataset, so the values of WQI are calculated based on the physicochemical parameters of pH, dissolved oxygen, turbidity, and temperature through a weighted summation approach. WQI can be formulated as follows.

$$WQI = \frac{\sum_{i=I}^{n} w_i q_i}{\sum_{i=I}^{n} w_i} \tag{2}$$

Where w_i is the weight that has been placed on the parameter and q_i s the quality rating of that particular parameter.

Graph Construction

The monitoring network is represented as a graph in order to capture space relationships between monitoring stations. The representation, in this case, uses each monitoring station as a node, and the spatial dependence between the monitors, e.g., the geographic distance or the hydrological connectedness, as the edges.

The graph structure is as given below in equation (3):

$$G = (V, E) \tag{3}$$

Where V represents the collection of monitoring stations, and E represents the spatial relationship between the monitoring stations.

A node has a feature vector that consists of the water quality parameters measured by the station over time. Spatial correlations of various locations can be well represented by modeling monitoring stations as a graph. The model is able to learn spatial interactions with an approach that is frequently ignored in conventional tabular machine learning techniques.

Spatial Feature Learning Using Graph Attention Networks

After creating the graph representation, spatial features are obtained with the help of Graph Attention Networks. GAT presents an attention system that enables the model to assign various weights to the importance of neighboring nodes.

The attention coefficient between node i and node j is determined by equations (4) - (8).

$$\alpha_{ij} = \frac{\exp(\text{LeakyReLU}(a^T[Wh_i || Wh_j]))}{\sum_{k \in N_i} \exp(\text{LeakyReLU}(a^T[Wh_i || Wh_k]))} \tag{4}$$

Where $h_i$and $h_j$represent node feature vectors, W is a learnable weight matrix, a represents the attention vector, and N_i represents the adjacent nodes of node i.

The node depiction is updated using these attention coefficients as follows.

$$h_i^{'} = \sigma\left(\sum_{j\in N_i} \alpha_{ij} W h_j\right) \quad (5)$$

Where σrepresents the activation function. In this process, the model acquires the knowledge of which monitoring stations are more spatially influential to each other.

Temporal Modeling Using Temporal Fusion Transformer

Once spatial embeddings are produced with the help of the Graph Attention Network, Temporal Fusion Transformer is used to model temporal dependencies in water quality data. Transformer-based models can learn time-series data with attention and learn long-term relationships among time-series data.

The self-attention mechanism that is used in transformer-based frameworks is defined as:

$$Attention(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (6)$$

Where Q, K, and V represent query, key and value matrices of input data, and d_k is the dimensionality of key vectors.

The TFT architecture combines attention modules and gating mechanisms, as well as feature selection networks. These components enable the model to capture temporal trends, seasonal changes, and nonlinear relations among environmental parameters. The hybrid GAT-TFT architecture is depicted in fig. 2.

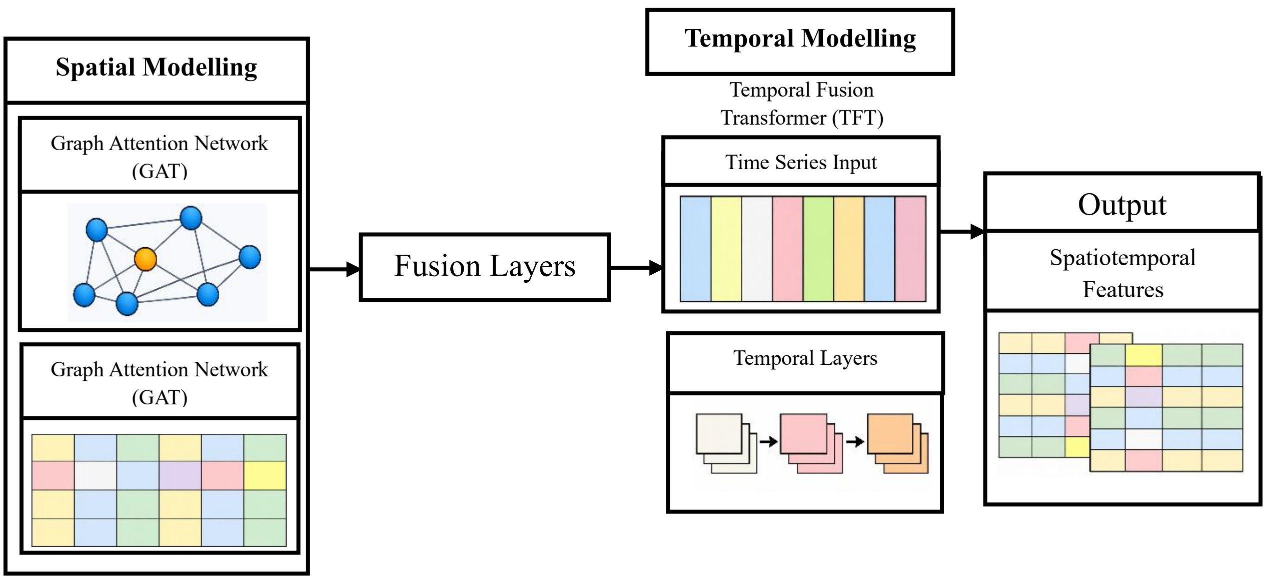

Fig. 2: Hybrid GAT–TFT Architecture for Spatiotemporal Feature Learning

WQI Prediction & Early Warning System

Once spatial and temporal features have been learned, the resultant representations are sent to a fully connected prediction layer. This layer determines future values of WQI using past water quality observations.

The loss function used to train the forecasting model is the Mean Squared Error, and it is defined as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (7)$$

Where y_i is the observed value of WQI and y_i is the estimated value.

In order to facilitate proactive environmental monitoring, the projected WQI values are compared with predetermined water quality criteria. An early warning alert is provided when the predicted value lies in either the bad or dangerous range of multiple observations. The environmental authorities are able to react as fast as possible and implement proactive actions before the quality of water gets out of control.

Explainability Using SHAP

Even though deep learning models might give correct predictions, they are not always transparent. To resolve this problem, SHAP (SHapley Additive Explanations) is applied in the interpretation of the prediction results.

The SHAP value for feature i can be expressed as

$$\phi_i = \sum\nolimits_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F|-|S|-1)!}{|F|!} [f(S \cup \{i\}) - f(S)] \quad (8)$$

Where F is the complete collection of input features, and S is subsets of features. The contribution that each parameter of water quality makes towards the predicted WQI can be identified by examining the SHAP values.

Algorithm

Algorithm 1 shows the entire process of the suggested hybrid GAT-TFT framework, which involves preprocessing, spatial--temporal modeling, prediction, and explainability, are provided in Algorithm 1.

Algorithm 1: Hybrid GAT–TFT Framework for WQI Prediction and Early Warning

Input: Water quality dataset D (pH,DO,turbidity,temperature,conductivity,salinity,BOD)

Output: Predicted WQI ($\hat{y}_t$), early warning alerts, feature explanations (ϕ_i)

1. Preprocessing: Interpolate missing values, remove outliers, and apply min-max scaling:
2. $x_{\text{norm}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$
3. Target Formulation: Compute baseline WQI via weighted aggregation if ground truth is missing.
4. Spatial Modeling: Construct graph $G = (V, E)$ from monitoring stations. For each node $i \in V$, compute attention coefficients α_{ij} and update representations h'_i.
5. Temporal Modeling: Apply self-attention mechanisms to capture time-series dependencies:
6. $\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$
7. Training & Prediction: Train using MSE loss ($\mathscr{L}_{\text{MSE}} = \frac{1}{N}\sum(y_t - \hat{y}_t)^2$) to predict future WQI ($\hat{y}_t$).
8. Classification: Categorize WQI:
9. WQI $\geq 70 \Longrightarrow$ Good

a. $50 \leq$ WQI $< 70 \Longrightarrow$ Poor

b. WQI $< 50 \Longrightarrow$ Hazardous

10. Alert Generation: Set Alert $= 1$ if WQI$(t - i) < 70$ for all $i \in [0, k - 1]$; otherwise Alert $= 0$.
11. Explainability: Compute SHAP values (ϕ_i) to evaluate feature contributions.
12. Return: Predicted WQI, alerts, and explanations.

Experimental Setup

In order to test the efficiency of the proposed Hybrid GAT-TFT framework, a set of experiments was implemented with practical data on water quality monitoring. The experimental design will be used to test the capability of the proposed model to understand both the spatial relationships among monitoring stations and the temporal dependence of environmental observations. This part explains the data set employed, data preparation process, model specification, evaluation criteria, and baseline models which are employed to compare performance.

Dataset and Data Preparation

This study used data accessed at the Water Quality Portal (WQP) on Data.gov. WQP is an open-source ecological database that consolidates data on aquatic monitoring entries gathered by various agencies, such as the U.S. Environmental Protection Agency (EPA), the U.S. Geological Survey (USGS), and other national and regional agencies. The data include multi-site measurements at various aquatic systems, including physicochemical indicators such as pH, dissolved oxygen, turbidity, thermal levels, conductivity, salinity, and biochemical oxygen demand (BOD). These measures are the key variables used to measure aquatic health and derive the Water Quality Index (WQI). The dataset was preprocessed before being used to train the model to guarantee reliability and consistency. Interpolation techniques were used to process missing values and statistical filtering was used to remove abnormal measurements that occurred due to sensor noise.

The processed dataset was then subjected to the min-max normalization process as (1) in order to normalize the data to the range of [0, 1], which enhances optimization and predictability of the model. WQI was not directly provided in the dataset and therefore was calculated by applying the weighted aggregation formula in (2) and taken as the target variable. In order to model spatial dependencies, the monitoring network was modeled as a graph G (V, E) in which the monitoring stations are nodes and the geographic distance is used as a spatial relationship as an edge, as in (3). The data were subsequently sorted into sequential windows of time to allow temporal learning, which were fed to the Temporal Fusion Transformer. Lastly, the data was divided into training (70%), validation (15%), and testing (15) sections. Optimization of model weights was performed on the training segment, refinement of configuration and generalization tracking on the validation data, and eventual assessment of efficacy was done on the testing subset. Table 1 gives the descriptive statistics of the datasets used during the experiment phase.

Dataset link: https://catalog.data.gov/dataset/water-quality-data-1367b

Table 1: Dataset Statistics Used in the Experiments

Dataset Property	Value
Data source	Water Quality Portal (WQP)
Monitoring stations	Bay (multiple station IDs present in dataset)
Total Observation	2500 samples
Input parameters	pH, DO, turbidity, temperature, salinity, BOD
Target variable	Water Quality Index (WQI)
Training set	70%
Validation set	15%
Testing set	15%

Model Configuration

The offered forecasting model combines both Graph Attention Networks (GAT) and Temporal Fusion Transformers (TFT) in the detection of multidimensional patterns in time and space within aquatic sensor data. Graph Attention Networks were employed in the spatial modeling stage to extract spatial features using the monitoring network. GAT layers calculate attention between adjacent nodes with (4) and update node representations with (5). Several attention heads were adopted to record different spatial interactions among monitoring stations. The GAT-generated spatial embeddings were then fed to the temporal modeling of the Temporal Fusion Transformer. To learn the long-term dependencies and time-varying patterns of multivariate time-series data, the TFT architecture employs the self-attention mechanism as in (6). It also has gating layers and selection networks that are variable and enable the model to automatically determine important environmental variables during training. The last prediction layer involves fully connected neural network layers that calculate future WQI values. Adam algorithm was applied to optimize the network weights and the architecture was trained with a series of iterations until it reached a steady performance. Table 2 provides the model hyperparameters that were used in the experiments.

Table 2: Model Hyperparameters Used in the Experiments

Parameter	Value
Learning rate	0.001
Batch size	32
Epochs	100
GAT attention heads	4
Hidden units	128
Dropout rate	0.2
Optimizer	Adam

Evaluation Metrics

Three common regression measurements were used to evaluate the predictive accuracy of the proposed framework: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the R-squared (R^2) are represented in equation (9) - (19).

Mean Absolute Error (MAE)

MAE is the mean value of the absolute deviations between the predicted and observed WQI values, which are formulated as.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i | \quad (9)$$

Where y represents the observed value of WQI and $\hat{y}_i$ is the predicted value.

Root Mean Squared Error (RMSE)

RMSE is the square root of the mean squared deviations of the forecasted and the actual observations and is defined as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (10)$$

This measure puts greater punishment on big prediction errors.

Coefficient of Determination (R²)

The coefficient of determination measures the ability of the outcomes being forecasted to explain the variation in the observed outcomes being measured:

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \dot{y})^2} \quad (11)$$

Where $\dot{y}$ represents the mean size of the observed WQI measurements.

Early Warning Evaluation Metrics

The proactive alert system is applied to assess the effectiveness of the framework with the help of the categorization metrics based on the confusion matrix. Predicted values of WQI are transformed into classes (Good, Poor, Hazardous) according to pre-defined values.

The measures used to make the assessment include Accuracy, Precision, Recall, and the F1-score:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}, \quad (12)$$

$$Precision = \frac{TP}{TP+FP} \quad (13)$$

$$Recall = \frac{TP}{TP+FN}, \quad (14)$$

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (15)$$

The multi-label categorization is applied in the context of the multi-label averaging procedure:

$$Precision_{macro} = \frac{1}{C}\sum_{i=1}^{C} Precisio\,n_i, \quad (16)$$

$$Recall_{macro} = \frac{1}{C}\sum_{i=1}^{C} Recal\,l_i \quad (17)$$

$$F1_{macro} = \frac{1}{C}\sum_{i=1}^{C} F\,1_i \quad (18)$$

Also, the False Alarm Rate (FAR) is calculated as:

$$FAR = \frac{FP}{FP+TN} \quad (19)$$

A reduced FAR will mean fewer false alarms and better reliability of the system.

Baseline Models for Comparison

In order to show the effectiveness of the developed hybrid GAT-TFT architecture, its results were compared to a number of standard benchmark paradigms commonly used in studies of aquatic quality forecasting. Such candidates are Artificial Neural Networks (ANN), Long Short-Term Memory networks (LSTM), and combined deep learning setups. All the reference models were trained on the same dataset and assessment criterion to enable a fair comparison. Results of these systems were used as benchmarks to compare the benefits of using spatial feature extraction with Graph Attention Networks and temporal sequence modeling with Temporal Fusion Transformers.

Results and Discussion

This part shows the empirical results of the proposed Hybrid Graph Attention Network and Temporal Fusion Transformer (GAT–TFT) architecture. Regression measures, which are used to assess the efficiency of the proposed framework, are outlined in (9), one of them being Mean Absolute Error (MAE), another one is the Root Mean Squared

Error (RMSE), and the value of the R-squared ($R^2). These findings are compared with some reference paradigms to illustrate how well the developed strategy can estimate the Water Quality Index (WQI).

Prediction Performance

The introduced architecture forecasting potential was tested on the evaluation dataset. The figures of the estimated WQI in the framework were compared with the ground-truth WQI values of the data subset. fig. 3 shows the relationship between the real WQI values and the expected levels of the WQI that would be generated by the Hybrid GAT-TFT system. The findings show that the predicted values are a close reflection of the trend of the measured WQI data. This has ensured that the setup is effective in capturing the temporal patterns in the aquatic information and produces accurate estimations.

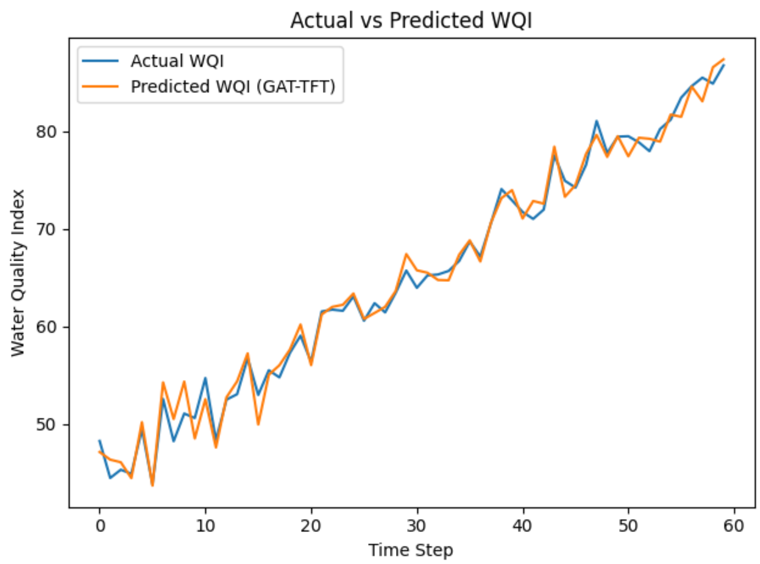

Fig. 3: Comparison between Actual and Predicted WQI Values by Hybrid GAT–TFT Framework

The excellent fit between the predicted and ground-truth values shows that the presented framework is capable of capturing rather complex and nonlinear relationships between the environmental factors and the Water Quality Index.

Model Performance Comparison

In order to promote a closer analysis of the effectiveness of the proposed architecture, the outcomes of the proposed architecture were compared with various benchmark paradigms that are generally used in the context of aquatic quality estimation. These include Artificial Neural Networks (ANN), Long Short-Term Memory networks (LSTM), combined LSTM models, and Graph Neural Network (GNN)-based frameworks.

Table 3: Performance Comparison of Prediction Models

Model	MAE	RMSE	R^2
ANN	0.084	0.116	0.89
LSTM	0.071	0.103	0.91
Hybrid LSTM	0.065	0.097	0.93
GNN-based Model	0.061	0.093	0.94
Proposed GAT–TFT	0.048	0.079	0.96

Table 3 demonstrates that the proposed Hybrid GATT-TFT model is the lowest in prediction errors compared to the model. The values of MAE and RMSE are much lower, whereas the value of R^2 is larger than the basic models. This comparison of the prediction errors of various models as measured by the Mean Absolute Error criterion is depicted in fig. 4.

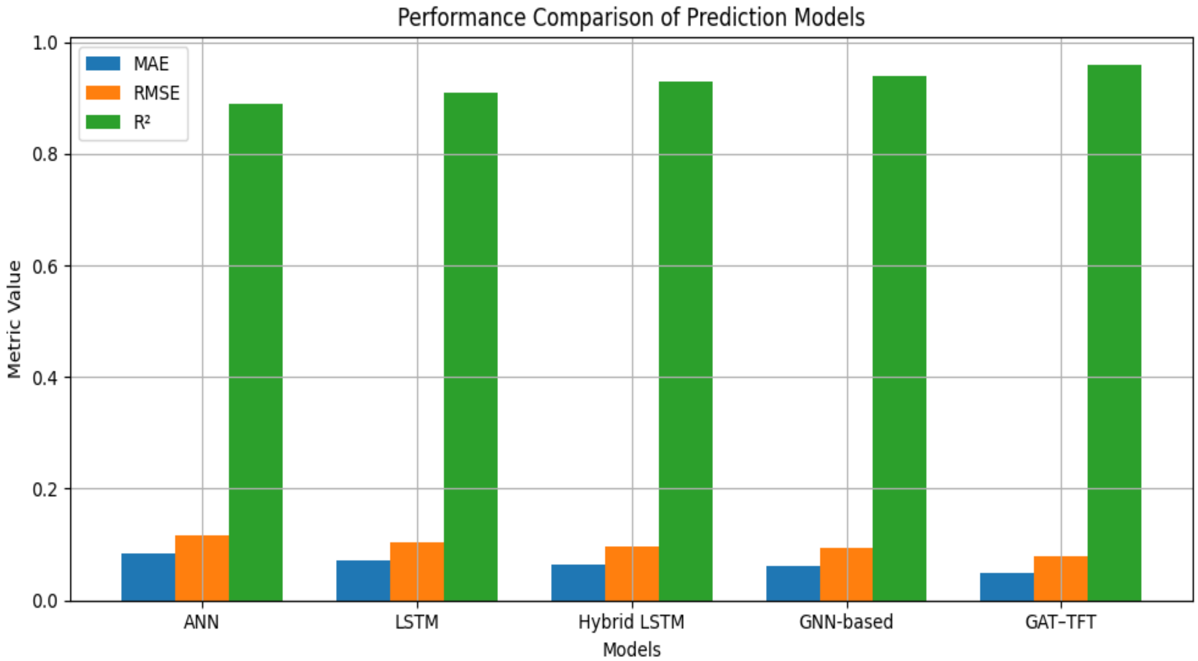

Fig. 4: Performance Comparison of Different Prediction Models Based on MAE

The increase in the accuracy of prediction is justified by the fact that the proposed framework has the capacity to model spatial relations among the monitoring stations and temporal relations among the environmental observations.

Spatial and Temporal Learning Analysis

The combination of the spatial and the temporal learning mechanisms can be attributed to the superior performance of the proposed model. The Graph Attention Network module represents spatial correlations between monitoring stations with attention coefficients as given in (4). These coefficients enable the model to attach dynamically some importance to adjacent monitoring stations. Consequently, the stations with similar environmental conditions influence more during prediction. Temporal Fusion Transformer component captures the temporal relationship within the dataset through self-attention mechanism as defined in (6). This will enable the model to reflect long-term relationships, seasonal relationships, and nonlinear relationships among water quality parameters. The Hybrid GAT-TFT framework offers a complex spatiotemporal water quality data representation by integrating the two mechanisms.

Early Warning System Analysis

WQI Threshold Classification

Besides forecasting the values of Water Quality Index (WQI), the proposed Hybrid GAT-TFT framework also includes an early warning mechanism to detect the possible water quality degradation beforehand. The estimated WQI values are then reduced into discrete levels of water quality using a set of threshold values in order to make actionable decisions.

Table 4: WQI Thresholds for Early Warning

WQI Range	Category	Alert Level
$WQI \geq 70$	Good	No Alert
$50 \leq WQI < 70$	Poor	Warning Alert
$WQI < 50$	Hazardous	Critical Alert

Table 4 shows the categories of Good and Poor and Hazardous describe conditions of risk of pollution that need to be monitored and mitigated; the former is safe water and the latter is the situation of growing levels of pollution risk.

Alert Generation Mechanism

The early warning system generates alerts based on predicted WQI values combined with a temporal persistence condition to ensure robustness. Let WQI_tdenote the predicted WQI at time step t. An alert is triggered only when the predicted WQI remains below the safe threshold for a predefined number of consecutive time steps. The alert condition is defined in equation (20):

$$Alert_t = \begin{cases} 1, & \text{if } WQI_{t-i} < 70 \forall i \in [0, k-1] \\ 0, & \text{otherwise} \end{cases} \quad (20)$$

Where k represents the persistence parameter, which is set to k=3 in this study. This formulation will guarantee that alerts are only triggered when the deterioration of water quality remains overtime and not because of temporary changes.

Temporal Consistency for False Alarm Reduction

The environmental monitoring data are usually noisy and subject to short-term fluctuations due to sensor malfunctions or short-term environmental fluctuations. Persistence parameter k is introduced to offer time consistency; this is needed to minimize false alarms. The system makes the use of several threshold violations in succession, effectively removing any isolated anomalies, and allows the alerts to reflect the long-term trends in pollution. This greatly enhances the level of reliability and strength of the early warning system when used in real-life situations.

Early Warning Detection Visualization

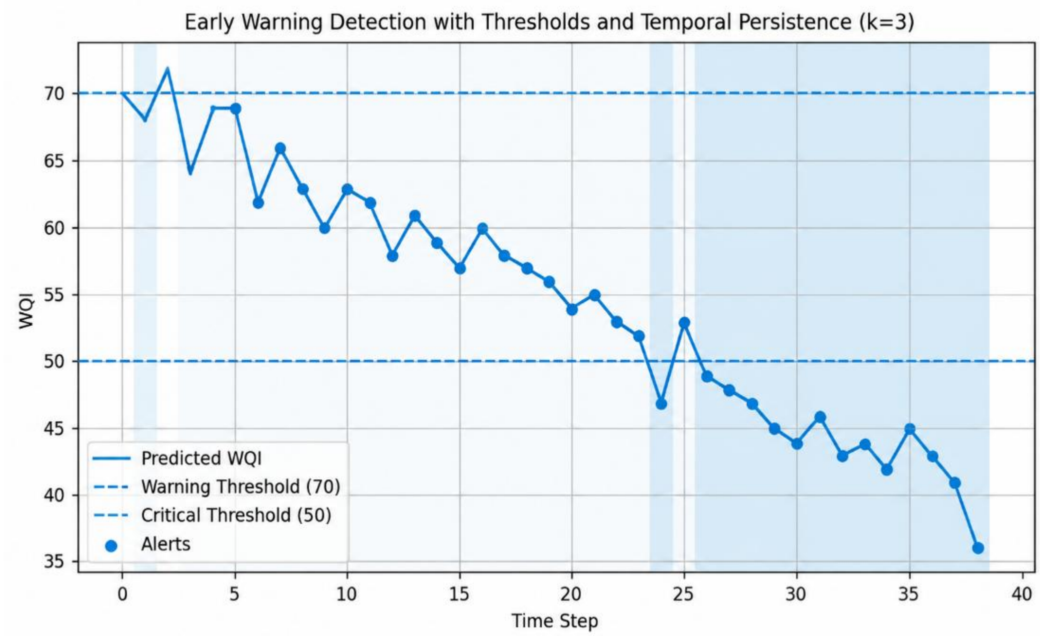

Fig. 5: Early Warning Detection Based on Predicted WQI Values and Predefined Threshold Levels

The process of early warning detection is shown in fig. 5. The predicted WQI values are plotted against time and the line of threshold (WQI=70). The highlighted areas denote the cases when the predicted values of the WQI are lower than the threshold over several time steps, which provokes warning or critical notifications. This shows how well the temporal filtering mechanism can identify long-term pollution incidences.

Performance Evaluation of Early Warning System

In order to make a quantitative assessment of the effectiveness of the early warning mechanism, the problem of prediction is re-formulated in the form of a classification task on the basis of WQI categories. Standard categorization indicators are used to measure the effectiveness of the framework in general.

Table 5: Early Warning System Performance

Metric	Value
Accuracy	0.92
Precision	0.89
Recall	0.94
F1-score	0.91
False Alarm Rate (FAR)	0.08

Table 5 represents the findings show that the proposed framework has better precision and sensitivity and proves to be competent in detecting aquatic degradation and reliable in raising notifications.

False Alarm Analysis

The rate of false alarms generated by the system is assessed through the False Alarm Rate (FAR) which is defined as in equation (21):

$$FAR = \frac{FP}{FP+TN} \tag{21}$$

Where FP represents false positives (incorrect notifications) and TN represents true negatives (correct non-alerts).

The FAR value of 0.08 is a low value that means that the system is efficient in reducing the number of unnecessary alerts and is highly sensitive to detecting. It is especially valuable in environmental monitoring where false alarms are too frequent to be trusted and it may result in poor resource use.

Confusion Matrix Analysis

To enable a more fine-grained analysis of categorization effectiveness, a confusion matrix is created in order to examine the prediction accuracy of the various WQI classes.

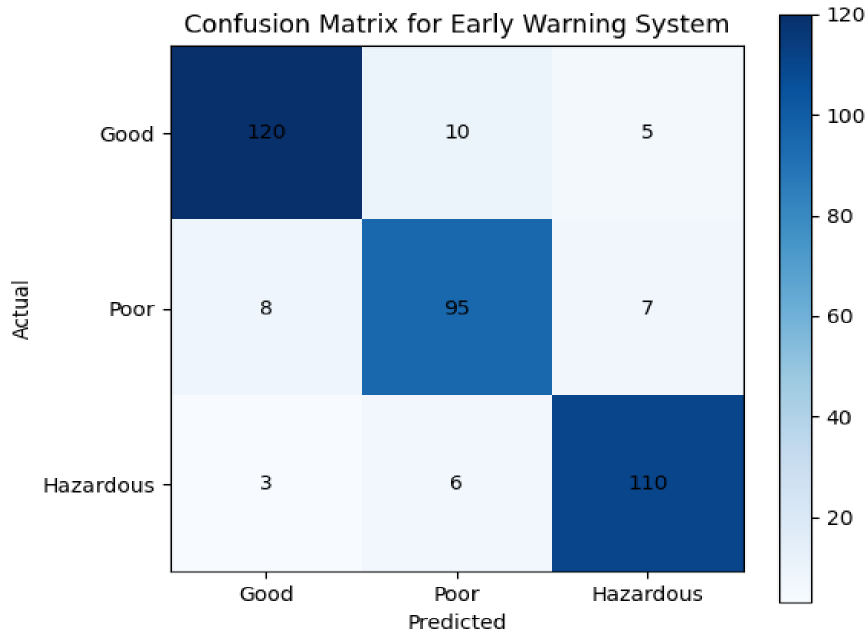

Fig. 6: Confusion Matrix for Early Warning System

In the confusion matrix, it can be seen that most of the estimations have been correctly categorized as the fig. 6 in the main diagonal are dominating. The misclassifications are small and occur mainly between the adjacent categories, as it should be expected since the threshold boundaries overlap.

Model Explainability

To improve model interpretability, SHAP explainability was used to identify the effect of each specific environmental parameter on the prediction results. The quality of attribute significance analysis, which is offered with the help of SHAP-generated explanations, is presented in fig. 7.

The findings show that the most influential parameters when predicting water quality include dissolved oxygen, turbidity and pH. It was established that higher turbidity levels had adverse effects on water quality, whereas an increased amount of dissolved oxygen had an overall positive relationship with the value of WQI. This explainability analysis allows the model to give useful insights on the environmental factors that have led to deterioration of water quality. The insights will assist environmental specialists to learn about sources of pollution and adopt proper mitigation measures.

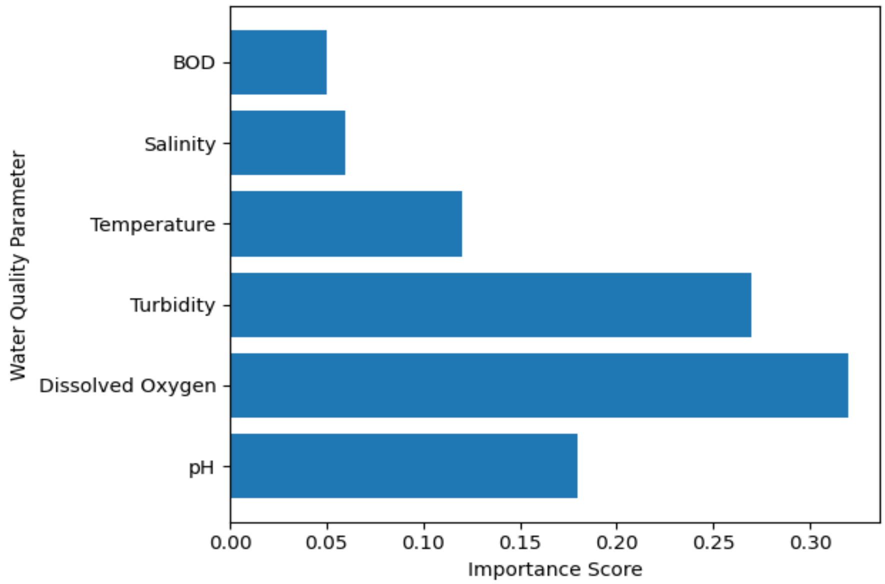

Fig. 7: Feature Importance for WQI Prediction

Conclusion

In conclusion, the above-mentioned research paper has successfully developed and validated a novel, explainable spatiotemporal deep learning framework which incorporates Graph Attention Networks (GAT) and Temporal Fusion Transformers (TFT) to make accurate predictions of the Water Quality Index (WQI) and generate early warnings for aquatic environments. Due to its departure from standard tabular-based methodologies, which enable modeling spatial relationships between the sensors and intricate temporal relations in ecological data, the suggested framework has outperformed traditional benchmarks in terms of predictive effectiveness by attaining an MAE of 0.048, RMSE of 0.079, and a high R^2 of 0.96. Moreover, the early warning function of the framework has proved to be highly effective with a classification accuracy of 0.92, recall of 0.94, and an impressively low False Alarm Rate (FAR) of 0.08. The method ensures effective filtering of temporary noise generated by sensors to avoid false alarms. In addition, the SHAP explainability ensured valuable insights into key environmental factors such as dissolved oxygen, turbidity, and pH that determine aquatic degradation. This smart monitoring system provides environmental authorities with a reliable mechanism for protecting their ecosystems. In future research, the framework could be extended through integration of real-time IoT sensors and other environmental factors to enhance large-scale, sustainable water resource management.

References

Cui, G., Zhao, Z., Wang, L., Zhang, H., & Yu, J. (2025). A Transformer-based water quality prediction model with mixed fractional Brownian features and multi-feature bottleneck transformation. *Applied Soft Computing*, 113860. https://doi.org/10.1016/j.asoc.2025.113860

Srinath, S., & Tuppad, P. (2025). Data driven multi-stage transformer-based framework for intelligent water quality monitoring. *Scientific Reports*, *15*(1), 41680. https://doi.org/10.1038/s41598-025-25527-x

Wan, H., Xiang, L., Cai, Y., Xie, Y., & Xu, R. (2025). Temporal and spatial feature extraction using graph neural networks for multi-point water quality prediction in river network areas. *Water Research*, *281*, 123561. https://doi.org/10.1016/j.watres.2025.123561

Wang, D., Zhang, C., Li, A., Guo, Y., Zhang, H., & Tan, C. (2025). Spatio-temporal analysis and prediction for raw water quality of drinking water source by improved RNN algorithm. *Journal of Water Process Engineering*, *71*, 107164. https://doi.org/10.1016/j.jwpe.2025.107164

Chukwuemeka, P., Imoni, O., Mogo, F. C., Eteh, D. R., William, T., Bamiekumo, B. P., ... & Ben-Koko, M. (2026). A hybrid AI model integrating LSTM, XGBoost, and K-means for interpretable prediction and clustering of water quality in data-scarce regions. *Discover Civil Engineering*, *3*(1), 24. https://doi.org/10.1007/s44290-026-00417-x

Ding, F., Zhang, W., Cao, S., Hao, S., Chen, L., Xie, X., ... & Jiang, M. (2023). Optimization of water quality index models using machine learning approaches. *Water research*, *243*, 120337. https://doi.org/10.1016/j.watres.2023.120337

Lv, J. Q., Yin, W. X., Xu, J. M., Cheng, H. Y., Li, Z. L., Yang, J. X., ... & Wang, H. C. (2025). Augmented machine learning for sewage quality assessment with limited data. *Environmental science and ecotechnology*, *23*, 100512. https://doi.org/10.1016/j.ese.2024.100512

Mahesh, N., Babu, J. J., Nithya, K., & Arunmozhi, S. A. (2024). Water quality prediction using LSTM with combined normalizer for efficient water management. *Desalination and Water Treatment*, *317*, 100183. https://doi.org/10.1016/j.dwt.2024.100183

Han, Z., Zhang, S., & He, L. (2025). Predicting and investigating water quality index by robust machine learning methods. *Journal of Environmental Management*, *381*, 125156. https://doi.org/10.1016/j.jenvman.2025.125156

Zheng, Y., Wei, J., Zhang, W., Zhang, Y., Zhang, T., & Zhou, Y. (2024). An ensemble model for accurate prediction of key water quality parameters in river based on deep learning methods. *Journal of Environmental Management*, *366*, 121932. https://doi.org/10.1016/j.jenvman.2024.121932

Ding, F., Hao, S., Jiang, M., Liu, H., Wang, J., Hao, B., ... & Zhang, Y. (2025). An improved graph neural network integrating indicator attention and spatio-temporal correlation for dissolved oxygen prediction. *Ecological Informatics*, *87*, 103126. https://doi.org/10.1016/j.ecoinf.2025.103126

Zha, Y., & Yang, Y. (2024). Innovative graph neural network approach for predicting soil heavy metal pollution in the Pearl River Basin, China. *Scientific Reports*, *14*(1), 16505. https://doi.org/10.1038/s41598-024-67175-7

Maqbool, S., Khan, M. I., Raza, A., Ali, M., Saddique, N., & Saddique, Q. (2026). A novel spatiotemporal and machine learning framework for sustainable groundwater monitoring and management. *Groundwater for Sustainable Development*, 101612. https://doi.org/10.1016/j.gsd.2026.101612

Nallakaruppan, M. K., Gangadevi, E., Shri, M. L., Balusamy, B., Bhattacharya, S., & Selvarajan, S. (2024). Reliable water quality prediction and parametric analysis using explainable AI models. *Scientific Reports*, *14*(1), 7520. https://doi.org/10.1038/s41598-024-56775-y

Zulkifli, S. N., Rahim, H. A., Noramli, N. A. S., Ali, M. S. M., Rahim, R. A., & Amirullah, I. (2026). Artificial neural network-based prediction of water quality for pollution assessment in freshwater ecosystems. *Water, Air, & Soil Pollution*, *237*(9), 522. https://doi.org/10.1007/s11270-026-09194-w

Perumal, B., Rajarethinam, N., Velusamy, A. D., & Sundramurthy, V. P. (2023). Water quality prediction based on hybrid deep learning algorithm. *Advances in civil engineering*, *2023*(1), 6644681. https://doi.org/10.1155/2023/6644681

Lin, F., Li, X., Su, Y., & Yan, J. (2025). Water quality prediction model based on improved long short-term memory neural network and empirical mode decomposition. *Discover Artificial Intelligence*, *5*(1), 199. https://doi.org/10.1007/s44163-025-00454-y

Liu, D., & Cheng, X. (2025). FNN-BiLSTM-Attention-DA: A hybrid fuzzy neural network and BiLSTM with multi-sensor information fusion for water quality monitoring and warning. *Alexandria Engineering Journal*, *124*, 624-639. https://doi.org/10.1016/j.aej.2025.04.011

Azma, A., Liu, Y., Azma, M., Saadat, M., Zhang, D., Cho, J., & Rezania, S. (2023). Hybrid machine learning models for prediction of daily dissolved oxygen. *Journal of Water Process Engineering*, *54*, 103957. https://doi.org/10.1016/j.jwpe.2023.103957

Yang, J. (2023). Predicting water quality through daily concentration of dissolved oxygen using improved artificial intelligence. *Scientific Reports*, *13*(1), 20370. https://doi.org/10.1038/s41598-023-47060-5