International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Temporal Fusion Transformer: A Deep Learning Approach for Modeling and Forecasting River Water Quality Index

Jitha P. Nair[1*], Vijaya M. S.[2]

Submitted: 26/05/2023 **Revised**: 06/07/2023 **Accepted**: 23/07/2023

Abstract: Water quality is a major factor when it comes to human and environmental health. The WQI is a key performance indicator for water management effectiveness. Water quality changes over time due to several seasonal attributes and physiochemical properties. As the seasons change at each site, the weather records are transformed into time series data, and the values of the physiochemical parameters shift accordingly. This paper introduces a novel temporal fusion transformer architecture for modelling and forecasting river water quality index. The WQI prediction model for the Bhavani River utilizes the temporal fusion transformer to incorporate temporal features from various scales of time series data obtained from monitoring stations. The performance results of the study are compared with other existing prediction models and demonstrated the effectiveness of the temporal fusion transformer approach for modelling and forecasting river water quality.

Keywords: *Deep Learning Architectures, Prediction, River Water Quality, Temporal Fusion Transformer, Time Series Data, Water Quality Index.*

1. Introduction

The ecosystem has been deteriorating and experiencing unanticipated impacts as a result of rising emissions and contaminants. As temperatures rise at unprecedented rates, the world's largest ice caps are melting at an incredible rate. The water data provides an excellent use case scenario, particularly when the provided data spans previous years from river monitoring stations. Water quality parameter levels fluctuate over time and have significant environmental impacts. Maintaining high standards of water purity is essential for both the preservation of our ecosystem and the well-being of people. It is essential for drinking, irrigation, recreation, and aquatic habitat. Poor water quality negatively impacts human health, such as spreading waterborne diseases and harming aquatic ecosystems and wildlife. Additionally, poor water quality affects industries that rely on water resources, such as agriculture and power generation.

Preserving both human health and the environment requires consistent efforts to maintain high water quality standards. Regular monitoring and testing are crucial for identifying and addressing any potential issues. Effective treatment and management of pollutants like sewage and agricultural runoff contribute to enhancing water quality. The conservation of water resources and regulations on industrial and agricultural activities are vital for their protection.

[1]**Research Scholar, Department of Computer Science, PSGR Krishnammal College for Women, Peelamedu, Coimbatore, India*
[2]*Associate Professor, Department of Computer Science, PSGR Krishnammal College for Women, Peelamedu, Coimbatore, India*

The Autoregressive method uses past time-series data to create a regression equation for prediction, while the Autoregressive Moving Average model combines AR and Moving Average methods to capture the linear relationship between variables over time. The Autoregressive Integrated Moving Average odel was developed to handle non-stationary data by pre-processing it before modelling with ARMA. Although other prediction techniques such as Hammerstein auto-regressive [2], Kalman filter [3], and Gray forecast [1] method are also available, they are primarily suitable for ultra-short-term or single-step forecasting and may not be effective for long-term temperature predictions where long-term dependencies between variables weaken over time.

Diverse artificial intelligence prediction approaches are employed for time series data forecasting, incorporating classical machine learning methods as well as deep learning techniques. Machine learning methods such as support vector regression [4], random forest, and XG Boost [5] have been successfully used in many prediction tasks. Traditional machine learning algorithms are not capable of handling data in the time dimension due to the identical relevance of data at each location, which obstructs the extraction of useful knowledge, but with the emergence of deep learning, significant learning methods such as Long Short-Term Memory Neural Network [8], Temporal Convolutional Network, and Transformer have been enhanced and implemented to extract time series features more effectively. Long Short-Term Memory Neural Network (LSTM) is an improved RNN model that selectively extracts pertinent historical data via a gate structure. In contrast to other sequence modelling methods, the Temporal Convolutional Network exhibits superior performance in processing sequence information, owing to its exceptional parallel processing capacity, consistent gradient propagation, and more adaptable receptive field compared to Long Short-Term Memory (LSTM) networks. Additionally, attention mechanisms have been applied to time series prediction tasks, such as in the Transformer model. The Transformer model is capable of identifying the crucial part of the input for each instance through the attention weight magnitude, which improves the model's interpretability. It also excels in capturing long-term dependencies compared to the RNN model.

The Water Quality Index serves as a means to evaluate and communicate the overall water quality of a specific water body. It is a numerical index that combines data from various water quality parameters, such as pH, dissolved oxygen, and pollutant levels, into one value that reflects the overall water quality. The WQI is utilized to determine the suitability of water for various uses such as drinking, irrigation, and recreation. WQI prediction models are used to forecast the WQI of a water body based on measured values of its water quality parameters. These models are constructed using historical water quality data and environmental factors like weather and land use, to predict future water quality conditions.

This research aims to develop a TFT-WQI prediction model to improve the accuracy and reliability of the water quality forecasting model by utilizing historical data and incorporating temporal dependencies between data points. Time series data are collected from sampling stations of the Bhavani River for 5 years 2016-2020. The dataset is prepared by carrying out preprocessing tasks like normalization, and feature selection, and named as WQI-SA dataset. The WQI prediction models are built using a sophisticated deep neural architecture, temporal fusion transformer, for forecasting the river water quality index.

2. Literature Review

Ensuring the safety and availability of clean water resources is heavily dependent on accurately predicting water quality, making water quality prediction a critical task. Over the years, researchers have proposed several models and techniques to predict water quality indicators based on various environmental factors. In recent years, deep learning-based models, such as the Temporal Fusion Transformer (TFT), have shown promising results in WQI prediction. This literature review aims to analyze the existing models on water quality prediction and evaluate the effectiveness of the TFT model in WQI prediction. In this paper, we will briefly discuss some of the important works in this field and provide insights into the potential of deep learning-based models for WQI prediction.

Abhay Srivastava and Alberto Cano [13] conducted a study to explore various deep learning approaches for analyzing and forecasting pH levels, including LSTM, GRU, RNN, and TFT models, to determine the most accurate algorithm to forecast the pH level. Widely monitored and learned the pH level in oceans to preserve the health of aquatic ecosystems. Their analysis revealed that the TFT model outperformed other deep learning approaches in accurately predicting pH levels. The researchers utilized the TFT architecture to forecast pH anomalies and determine the significance of the predicted data.

Theyazn Aldhyani, Al-Yaari, et. al [16], had developed advanced artificial intelligence methods to predict water quality index and water quality classification. Artificial neural network models, including the nonlinear autoregressive neural network and the long short-term memory deep learning algorithm, had been created for the WQI prediction. The WQC forecasting employed three machine learning algorithms: support vector machine, nearest neighbour (K-NN), and Naive Bayes. The utilised dataset consisted of seven significant parameters, and the created models was tested using several statistical metrics. The outcomes had shown that the suggested models could reliably forecast WQI and categorize water quality based on greater resilience. The NARNET model fared marginally better than the LSTM model for the prediction of WQI values, while the SVM algorithm had the highest prediction accuracy (97.01%) for WQC values.

Zhenbo Li, Fang Peng, et al. [12] proposed a hybrid model to enhance the precision of DO prediction in aquaculture. The model is based on a sparse auto-encoder and long-short-term memory network. The SAE pre-trained hidden layer data comprises deep latent properties of water quality, which is then fed into the LSTM to progress the forecast accuracy. The experimental outcomes yielded that the SAE-LSTM hybrid model outperformed LSTM by reducing MSE in the prediction stages. The hybrid model also outperformed SAE-BPNN by 87.7%, 91.0%, and 90.0%, indicating that it was more accurate.

Lim, B., Arik, S. et.al [11], had presented that multi-horizon forecasting issues frequently comprised of a complicated mixture of inputs such as static, i.e., time-invariant variables, known future inputs and other exogenous time series that had only been seen historically, with no previous knowledge. They had presented several deep learning models for multi-step prediction but they typically consisted of black-box models that did not account for the entire range of inputs available in typical cases. In the article, they introduced the Temporal Fusion Transformer, a novel attention-based architecture that combined high-performance multi-horizon forecasting with interpretable insights into temporal dynamics. The Temporal Fusion Transformer (TFT) model utilized recurrent layers for local processing and interpretable self-attention layers to learn long-term dependencies and temporal relationships at different scales. The model had particular components to select appropriate features and gating layers to suppress unnecessary components, which resulted in high performance across a wide range of operating conditions. On a variety of real-world datasets, it demonstrated considerable performance improvements over previous benchmarks and presented three practical applications of TFTs interpretability.

Yurong Yang et al. [14], proposed a CNN-LSTM with Attention (CLA) water quality prediction model to predict water quality attributes. The study was conducted on the Beilun Estuary and the developed water quality dataset was used to estimate pH and NH3-N. The missing data was filled using wavelet technique and linear interpolation was used to denoise the data. The CNN-LSTM hybrid model was effective in addressing nonlinear time series prediction issues, and the attention mechanism was capable of capturing longer time dependence. The experimental outcomes exhibited that the model outperformed others in providing stable predictions with varying time lags.

The literature review has shed light on the various models and techniques proposed for water quality prediction. It is evident that traditional statistical methods, such as ARIMA and regression models, have been widely used for WQI prediction, but they have limitations in capturing complex nonlinear relationships. Recent studies have shown that deep learning-based models, such as the Temporal Fusion Transformer (TFT), have outperformed traditional models and achieved remarkable results in WQI prediction. TFT's ability to handle temporal dependencies and capture long-term patterns makes it a promising model for predicting WQIs. However, further research is needed to optimize the model’s hyperparameters and address issues related to data preprocessing and feature engineering. The literature review suggests that deep learning-based models significantly improve the accuracy and reliability of WQI prediction, and it opens up exciting opportunities for further research in this field.

3. Temporal Fusion Transformer

The Temporal Fusion Transformer is a modern method for time series analysis using deep learning techniques, which has proven to be effective and efficient in various applications. TFT is a multi-horizon model which incorporates a vast array of covariates into projections. The model accepts both static and dynamic variables, the effects of which are concealed from the user. The model is equipped with a temporal self-attention decoder that permits it to learn long-term patterns by considering the adaptability and architecture of the Temporal Fusion Transformer. They are a combination of transformer models, which are commonly used for natural language processing tasks, and temporal convolutional networks with time series data.

The TFT utilizes the transformer architecture and temporal fusion mechanism to effectively build forecasting models for future predictions. The transformer architecture allows TFT to handle large amounts of data and incorporate multiple data sources, such as sensor data and weather data, providing a more comprehensive prediction. The temporal fusion mechanism effectively combines the temporal information from different data sources to make accurate predictions. This principle of incorporating temporal information from multiple data sources allows TFT to account for temporal variability in accurate predictions. TFT is based on the transformer architecture, which is known for its ability to effectively process sequential data, this makes TFT a great model for time series data.

Temporal Fusion Transformers architectures have several advantages over other deep learning architectures when building prediction models. Compared to Recurrent Neural Networks (RNNs), TFTs have the ability to handle sequential data with much longer time steps than RNNs. TFT is an effective choice when building prediction models for tasks such as natural language processing and time series forecasting, due to its ability to handle sequences of varying lengths effectively, and its power in handling sequential data.

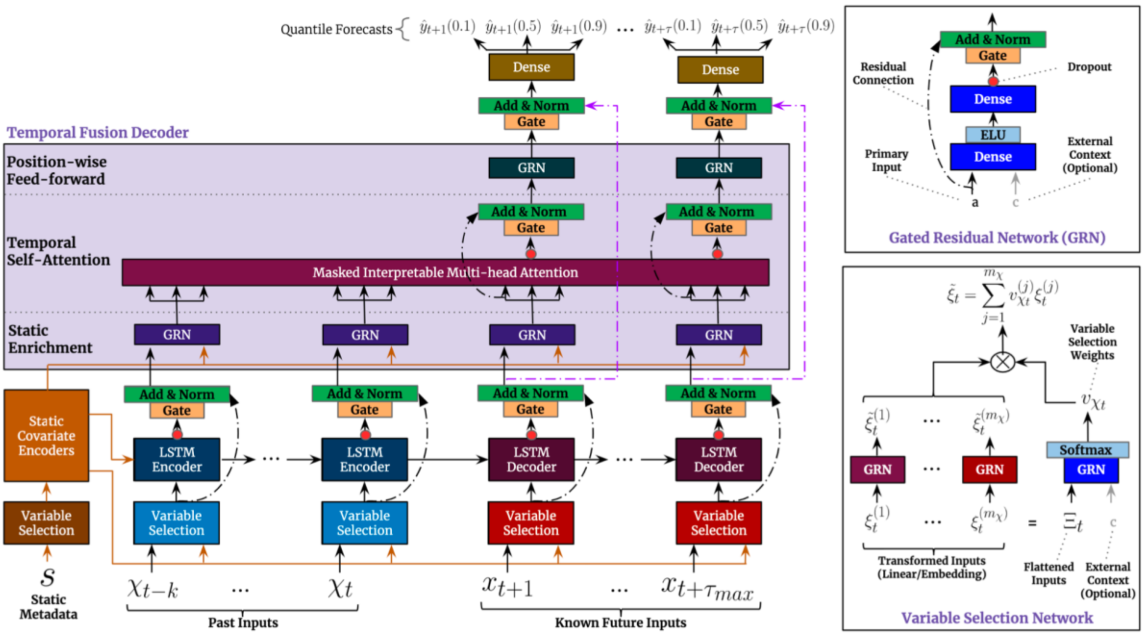

Fig.1. Framework for Temporal Fusion Transformer Design [15]

The TFT is a deep learning model that stands out from other models due to its unique building blocks, such as variable selection networks, gating mechanisms, prediction intervals, static covariate encoders, and temporal processing. The TFT's variable selection networks enable it to select only the most critical variables for data prediction, including both constant and time-varying variables. Gating mechanisms provide flexibility and avoid unnecessary design components, allowing the model to selectively produce nonlinear outcomes. Static covariate encoders use gated residual networks (GRNs) to produce weight vectors for encoding static covariates. Temporal processing is achieved through a sequence-to-sequence layer for local processing and a multi-head attention block for examining long-term dependencies. The TFT combines the LSTM encoder-decoder layer with other layers, such as self-attention layers, to improve model performance. Overall, the TFT is a powerful deep learning model that can learn both long-term and short-term associations between time-varying inputs, making it a valuable tool in data prediction and analysis.

The Multi-Head Attention (MHA) layer is a crucial element of the Transformer architecture, enabling the model to attend to multiple aspects of the input simultaneously. In the Temporal Fusion Transformer, the MHA layer is utilized to attend to different time steps of the input sequence, enabling the model to capture short and long-term trends. Comprising multiple sub-layers, each attending to a different aspect of the input, the MHA layer allows the model to interpret information differently, identifying complex input patterns and temporal correlations. The Gated Residual Network (GRN) is a key building block throughout TFT, consisting of ELU and GLU activation functions and two dense layers. The GLU is utilized to identify the most important features for prediction, and the activation functions assist the network in comprehending which input transformations are simple and which require more complex modelling. The TFT's GRN output goes through standard layer normalization and has a residual connection, potentially bypassing the input depending on its location, and static variables are used accordingly. Overall, these elements enable the TFT to identify patterns in the input that may not be immediately obvious and provide a powerful tool for data prediction and analysis.

The Temporal Fusion Transformer is a powerful new type of deep learning architecture that offers a number of advantages over variants of recurrent neural networks. Temporal Fusion Transformer enhances the interpretability of time series forecasting by identifying globally relevant variables, enduring temporal variations, and significant events for the prediction problem. TFTs can capture both long-term and short-term dependencies in the data, allowing for more accurate predictions. TFTs are able to capture complex relationships in the data that are difficult for LSTM, and GRU to learn. Hence TFT is chosen in this work for efficiently building a time series river water quality prediction model.

4. Materials and methods

The water quality index (WQI) of a river is a significant parameter used for assessing water quality since it offers a comprehensive evaluation of the overall water quality status. This study proposes a new method for modelling and forecasting river WQI using temporal fusion transformers (TFT). TFTs are a type of deep learning model to capture the temporal dependencies between time series data. The time series river water quality data was used for training TFT, and various hyperparameters were appropriately defined while creating the model. Various phases of building the WQI forecasting model are outlined below. The methodology employed to generate the WQI prediction model comprises the following phases: 4.1. Data Acquisition and Dataset Preparation 4.2. Exploratory data analysis and data preprocessing 4.3. Building the WQI prediction model 4.4. Validation and Model evaluation. Fig.2 illustrates the workflow of the proposed research.

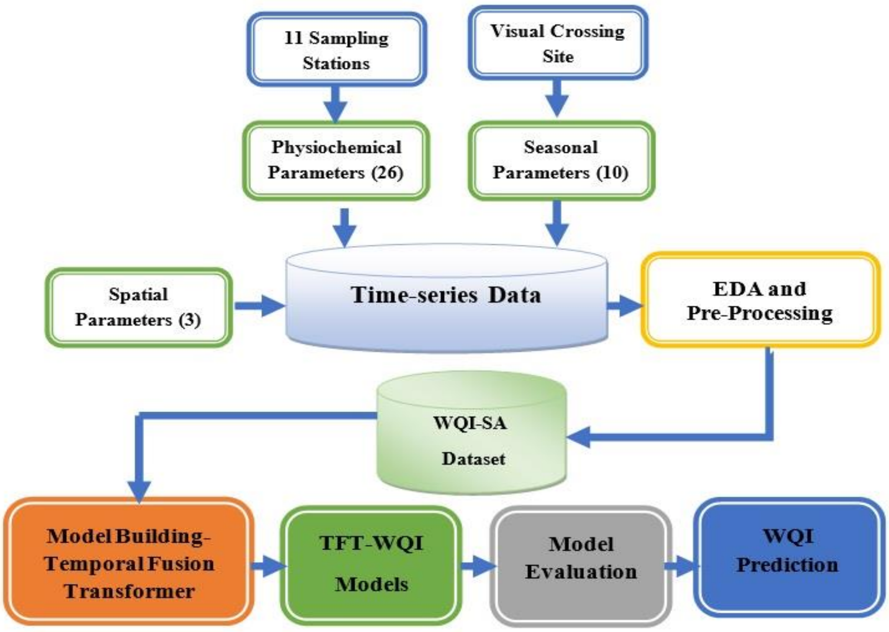

Fig. 2. Architecture of the Proposed TFT-Based WQI Prediction Model

4.1 Data Acquisition and Dataset Preparation

Real-time time series data were collected from sampling points along the Bhavani River for the period from 1st January 2016 to 31st December 2020. Twenty-six physiochemical water quality parameters measured daily at eleven sampling locations along the Bhavani River have been acquired. The seasonal parameters have been collected from the visual crossing site corresponding to the locations of the sampling stations.

Physiochemical Parameters

The evaluation of water quality heavily relies on physiochemical parameters, which offer essential insights into the physical, chemical, and biological properties of the water. Physical parameters include temperature, total suspended solids (TSS), turbidity, fixed dissolved solids (FDS), conductivity, and total dissolved solids (TDS). Conductivity measures the ability of water to conduct electricity, while turbidity indicates the cloudiness of water caused by suspended particles. Fixed solids in water refer to TSS and TDS residues remaining after heating to dryness.

The chemical parameters include pH, biological oxygen demand, ammonia, alkalinity, fluoride, chloride, biological oxygen demand, potassium, sulphate, dissolved oxygen, nitrogen, hardness, and chemical oxygen demand. A pH of 7 represents neutrality, below 7 indicates acidity, and above 7 indicates basicity. High ammonia levels in river water indicate increased pollution from dead plants and animals, algal growth, and faecal matter. Alkalinity is the sum of all soluble solids based on acid-neutralizing capacity, used in water softening calculations. Chloride concentration in freshwater is a sign of contamination, with sources including agricultural runoff, wastewater, and chloride-containing rock. High sulphate levels in natural water are due to magnesium or sodium sulphate deposits leaching. Nitrate levels in surface water can degrade water quality and may come from chemical fertilizers in farming activities. Hardness is characteristic of heavily mineralized waters. Dissolved oxygen (DO) is a major indicator of water pollution, with high concentrations signifying better water quality.

The biological water quality indicators are total coliform and faecal coliform, with the presence or absence of living organisms being useful indicators. The physiochemical parameters used for the research are added in Table 1 and are essential for determining the safety of water for drinking, irrigation, and other purposes.

Table 1. Physiochemical Parameters used in WQI Prediction

Parameters	**BIS Standard (Sn)**	**Parameters**	**BIS Standard (Sn)**
Temperature	28	pH	8.5
Turbidity	5	Ammonia	50
Conductivity	150	Alkalinity	200
TSS	300	Chloride	250
TDS	1000	Potassium	2.5
FDS	200	Sulphate	200
TC	100	Fluoride	1.5
FC	60	Hardness	100
BOD	3	DO	7.5
COD	10	Nitrate	0.503

Seasonal variations are an important factor in assessing water quality, as they reflect the changes in water quality caused by seasonal patterns and fluctuations. Seasonal parameters, including dew point, humidity, barometric pressure, precipitation, precipitation amount, wind speed, wind direction, cloud cover, and visibility, are key indicators of changes in weather and atmospheric conditions. The temperature of water affects the growth and survival of aquatic life, and changes in temperature can indicate changes in water quality. Precipitation affects water levels, influencing the movement and distribution of water and pollutants. Variations in water levels also reflect changes in water storage and usage during different seasons.

The data collected from eleven sampling stations and visual crossing sites from January 1st 2016 to December 31st 2020 and are converted into a time series dataset with parameters physicochemical, seasonal parameters, station ID and location data with 10,560 instances.

Computation of WQI

WQI is a measure for reporting the overall water quality of a specific location and can be used to identify areas that need improvement or to compare the water quality of different locations. It is based on the concentration of various chemical, physical, and biological parameters used to assess water's suitability for a specific use. The WQI can be computed using different methods, such as the Canadian Council of Ministers of the Environment (CCME) method, the National Sanitation Foundation (NSF) method, and the BIS Indian Standard. Here the WQI is determined based on Indian Standard for Drinking Water Specification (BIS 2004).

The following steps are followed to calculate WQI. First, weights are assigned to each water quality parameter based on their relative importance. The relative weight is calculated by dividing the weight of each parameter by its permissible limit (Si). Next, a quality of water rating (Qi) is assigned for each parameter based on its mean concentration value compared to the desirable limit as per the Indian drinking water standard (BIS 2004). The sub-index (SI) for each water quality parameter is then calculated by multiplying the relative weight by the quality of water rating. Finally, the WQI is calculated by summing the sub-index of each water quality parameter. The method of WQI calculation is presented in detail in Table 2.

Table 2. Computation of Water Quality Index

Parameters	BIS Standard (Si)	1/Si	K= ¹⁄∑¹⁄Si	Wi= K/Si	Ideal Value	Mean Value (Vi)	Qi= Vi/Si *100	SI= Wi*Qi
Temp	28	0.03	0.118	0.004	0	28	40	0.169
pH	8.5	0.11	0.118	0.013	7	7.3	85.88	1.202
Conductivity	150	0.006	0.118	0.0007	0	65	43.33	0.034
Hardness	100	0.01	0.118	0.0011	0	9	9	0.01
Sodium	200	0.005	0.118	0.0005	0	7	3.5	0.002
TSS	300	0.0034	0.118	0.0003	0	300	100	0.03
BOD	3	0.334	0.118	0.0396	0	2.3	76.67	3.04
Nitrate-N	0.503	1.988	0.118	0.236	0	0.902	179.3	42.41
TC	100	0.01	0.118	0.0011	0	60	60	0.071

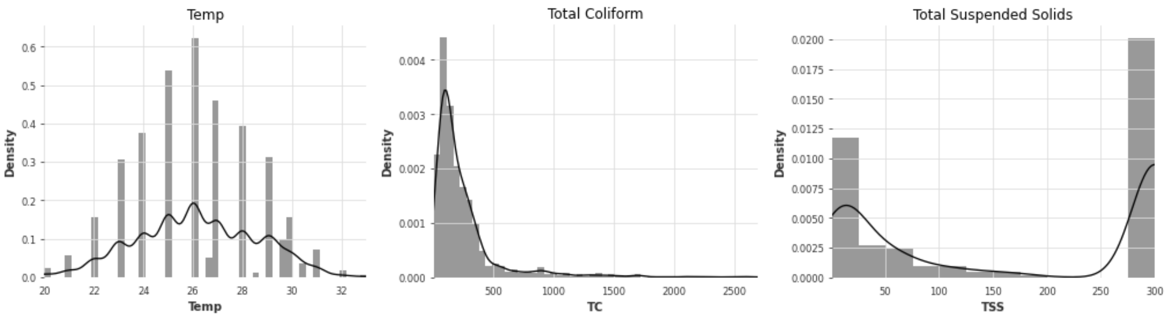

Fig.3. Distribution of Water Quality Parameters

The table 3a provides descriptive statistics of various water quality parameters measured in 10560 samples. The least value of conductivity is 6.4, the highest is 1207, the mean is 188.23, the median is 160.4, and the standard deviation is 127.09. The conductivity values are distributed widely, with a relatively high standard deviation. The mean value of 188.23 is higher than the median value of 160.4, indicating that the data are positively skewed. The least TSS value is 1, the highest is 300, and the mean value is 169.14, which means relatively high variation. The median value, is 300 and the standard deviation is 134.56, this indicates that there is a wide range of TSS values with a large variation from the mean. The lowest TDS value is 10 and the highest is 1175, which means a large variation in observation. The mean TDS value is 128.63 mg/L and the median is 115 mg/L. The standard deviation of the TDS values is 92.78 mg/L, indicating the TDS values with large variation and are positively skewed.

The results show that the lowest value of FDS is 0.02, the highest value is 772, the mean is 149.74, the median is 125, and the standard deviation is 100.72, with a relatively high standard deviation. The median value of 125 suggests that half of the samples had FDS concentration below 125 and a half had a concentration above 125. The standard deviation of 100.72 specifies that the FDS values in the samples varied widely from the mean value of 149.74.

The total coliform results show that the least value found is 8 and the supreme is 2691, which indicates high variability. The average TC value is 257.79 and the median value is 158. The standard deviation is 319.59, which indicates the level of variability in the results. The high standard deviation shows that the TC values are spread out over a large range. The results of FC show that the values range from a least of 10 to a supreme of 2186, with a relatively high variation in values. The mean value of FC is 117.83, with a median of 70 and a standard deviation of 178.4. The values of FC in the samples are spread out, with some samples having much higher values than others. The median value of 70 is lower than the mean value, which suggests that there was a higher concentration of low values compared to high values in the data set. Overall, the results indicate that there is some variability in the levels of FC. A low standard deviation denotes that the data points are clustered closely around the mean value for the given water quality parameters. The large standard deviation indicates that the data points are widely spread out from the mean, it is found for water quality parameters such as conductivity, TSS, TDS, FDS, TC and FC, as the values have large outliers these parameters need to be normalized.

Table 3a: Descriptive Statistics of Physiochemical Parameters

	Count	**Min**	**Max**	**Mean**	**Median**	**SD**
pH	10560	5.9	8.76	7.45	7.49	0.49
Conductivity	10560	6.4	1207	188.23	160.4	127.09
Turbidity	10560	1	332	7.39	2	20.98
PA	10560	0	26	0.7	0	2.31
TA	10560	1	804	67.87	63	43.76
Chloride	10560	0	215	19.27	14	17.86
COD	10560	0.12	24	7.78	4	5.53
TKN	10560	0	39	0.9	0.1	1.47
Ammonia	10560	0.21	5.39	0.54	0.25	0.49
Hardness	10560	4	298	75.2	67	48.87
Ca. hardness	10560	1	430.1	34.04	25	30.17
Mg. hardness	10560	0.62	110	20.59	15	18.09
Sulphate	10560	0	55	7.39	6	7.69
Sodium	10560	0	182	13.25	9	15.31
TSS	10560	1	300	169.14	300	134.56
TDS	10560	10	1175	128.63	115	92.78
FDS	10560	0.02	772	149.74	125	100.72
Phosphate	10560	0	1.5	0.2	0.11	0.19
Boron	10560	0	0.1	0.07	0.1	0.05
Potassium	10560	0.01	29	2.31	2	1.55
BOD	10560	0	6.5	1.35	1.13	0.9
Fluoride	10560	0	9.4	0.47	0.39	0.68
Nitrate-N	10560	0	11.42	0.71	0.54	0.88
TC	10560	8	2691	257.79	158	319.59
FC	10560	10	2186	117.83	70	178.4

The summary statistics of 10,560 observations of various seasonal variables are shown in Table 3b. The statistics show that precipitation, with the lowest value recorded as 0 and the highest value is 251 units. The mean value of precipitation is 9.2 units with a median of 1 unit and a standard deviation of 18.16 units, which indicates a large variation in values. It is concluded that the value of precipitation recorded is relatively low with a wide variation in the observations recorded. The mean and median being different suggests that the data is not evenly distributed, but rather has a skewed distribution.

The cloud cover data in the given statistics show that the smallest recorded cloud cover is 1.2, the highest is 99.9, and the average is 48.98, with a relatively high deviation among observations. The median is 48.7, and the standard deviation is 18.92. The cloud cover is a relatively wide range of values as compared to its mean value. The wind direction has a wide spread of values, as indicated by the relatively large standard deviation of 69.26. The median of 170.85 suggests that half of the wind direction observations are between 170.85. The mean of 154.7 indicates that the wind direction tends to be in the 154.7-degree direction, on average.

A large widespread of values are found for seasonal parameters such as precipitation, wind direction and cloud cover, and all other parameters are found to have small variability in observations.

Table 3b: Descriptive Statistics of Seasonal Parameters

	Count	**Min**	**Max**	**Mean**	**Median**	**SD**
Temp	10560	20	33	26.16	26	2.41
Dew	10560	3.3	24.7	20.09	21	2.78
Humidity	10560	28.44	97.27	70.46	72.18	10.32
Sea level pressure	10560	987.4	1020.4	1009.45	1009.2	2.61
Precipitation	10560	0	251	9.2	1	18.16
Precip cover	10560	0	100	7.04	4.17	13.42
Windspeed	10560	0.1	268.6	17.65	18.4	9.82
Wind dir	10560	1.2	337	154.7	170.85	69.26
Cloud cover	10560	1.2	99.9	48.98	48.7	18.92
Visibility	10560	2.2	10	5.52	5.4	0.98

The examination of the descriptive statistics of physiochemical and seasonal parameters is crucial in gaining a deeper understanding of their distribution and variability. This analysis enables the identification of patterns and relationships that is leveraged to improve WQI prediction. The results of exploratory data analysis prompt that, (i) it is essential to normalize the data and (ii) to clean the noise data by handling missing values and outlier analysis, enables to construct an efficient dataset for WQI prediction.

Preprocessing and Feature Selection

Preprocessing and feature selection plays a vital role in the analysis of WQI data and to produce meaningful results. The preprocessing phase involves cleaning and transforming the data, eliminating any missing values, outliers, and inconsistencies. In this research, the preprocessing tasks such as handling missing values, removal of outliers, and data normalization has been carried out. To normalize the attributes with high variation in their observations, the WQI-SA dataset is subjected to min-max normalization.

In prediction modelling, feature selection is a critical stage that involves the selection of parameters that have a significant impact on the prediction of the target variable. The Select K Best algorithm is employed here to identify the most important features in the prediction of the Water Quality Index. The Select K Best algorithm is a feature selection technique that aims to select a subset of the most informative features from a larger set of features. Once the features have been scored, the top K features with the highest scores are selected and used in further analysis or modelling. This helps to reduce the dimensionality of the data and increase the accuracy and efficiency of the analysis or modelling. The process revealed that conductivity is the most important feature, followed by ammonia and phosphate in computing WQI. On the other hand, boron and phenolphthalein alkalinity were ranked negatively, have limited impact and hence these parameters are removed from the dataset. The selected parameters are used to develop an efficient dataset to train TFT based WQI prediction model.

4.3 Building the WQI Prediction Model

The task of predicting the water quality index is approached as a regression problem and is tackled using temporal fusion transformer architectures. WQI prediction models are developed using temporal fusion transformer architecture by training the time series dataset.

TFT architecture designed for time series prediction tasks. TFT works by combining multiple layers of transformer blocks to capture long-term dependencies in the time series data. The input layer takes in the input WQI-SA time series data and converts it into a numerical representation. The encoder layers use self-attention mechanisms to process the input data and generate feature representations for each instance in the series and allow information from past and future time steps to be fused into the representation for each step. The decoder layers use self-attention and cross-attention mechanisms to generate predictions for future time steps based on the encoder's feature representations. Finally, the output layer generates the final WQI predictions for future time steps by applying a linear transformation to the output of the decoder. The number of layers, the size of each layer, and the hyperparameters of the model are tuned to achieve the best performance on a specific time series prediction task.

The TFT model uses a combination of two popular architectures for sequential data processing such as transformers and LSTM. The Transformer architecture allows for parallel computation and captures long-term dependencies in the data through self-attention mechanisms. The LSTM architecture is designed to handle sequential data and is capable of capturing short-term dependencies in the data. By combining these two architectures, TFT is able to effectively capture both long-term and short-term dependencies in time series data.

The model is trained using the specified number of epochs, and the learning rate is set to the specified value. The model uses the Quantile Regression likelihood function and is set to log information in the tensor board. The model is set to be randomly initialized and is forced to reset, which means any information about the previous model. The model is set to save checkpoints during training, which allows for quick and convenient recovery when training is interrupted.

Various hyperparameters such as the number of hidden layers, the number of neurons in each layer, and the learning rate are set before training. An optimizer is an algorithm used to minimize the loss function in a deep learning model and update the model's parameters. Here adam optimizer is used. A dense layer is a fully connected layer in a neural network, where each neuron is connected to every neuron in the previous layer. An epoch is one complete iteration over all the training data and the number of epochs to run during training is a hyperparameter. The batch size is the number of samples used in one iteration to update the model's parameters, with a larger batch size leading to faster convergence and a smaller batch size providing a more accurate solution. Dropout is a regularization technique commonly used in deep learning that randomly drops out a percentage of neurons in a neural network during training to prevent overfitting. Momentum is a technique used in optimization algorithms to speed up convergence by adding a fraction of the update vector from the previous time step to the current update vector. The learning rate is the step size at which the optimizer updates the model's parameters, with a larger learning rate leading to quicker convergence and a smaller learning rate leading to a more accurate solution.

A temporal fusion transformer is a variant of the transformer architecture designed for processing sequences of variable length. Some of the special hyperparameters used in TFT that differ from other deep learning architecture include attention windows, filter heads, value dimensions, and temporal encoder dimensions. Attention windows use a sliding window mechanism for attention computation, where the size of the window determines the range of context considered for each position in the sequence and helps to control the amount of context used for each prediction. The number of filter heads used for attention computation and their value is often in the range of [1, 16]. A larger number of filter heads help the model capture different types of dependencies between elements in the sequence. TFT uses a query, key, and value mechanism to compute attention scores and the dimension of the query, key, and value vectors are determined, which controls the model's capacity to capture dependencies in the data. The temporal encoder is used to project the input sequences into a higher-dimensional space and the dimension of the encoder's output helps to control the model's ability to capture complex relationships in the data. These special hyperparameters are important for TFT to perform well on sequence data and significantly impact the performance of the model.

Here, a particular TFT model is instantiated using the TFT Model class and is initialized with various hyperparameters such as input_length, output_length, hidden_layers, lstm_layers, num_attention_heads, dropout, batch_size, n_epochs, nr_epochs_val_period, likelihood, optimizers, random_state, reset, and checkpoints. The number of training iterations (EPOCHS), input size (INLEN), and expected features (FEAT) in the inputs are defined to optimize the model's performance. The model also features a self-attention mechanism with a specified number of heads, and defined encoder and decoder layers. The dimensions of the feedforward network, batch size during training, activation function, and learning rate are also defined to fine-tune the model's efficiency. The random seed used during training and the lower and upper bounds of the predictions, qL1, qL2, qL3, and qU1, qU2, qU3, respectively, further enhance the model's performance.

The WQI prediction model is built by training the temporal fusion transformer with WQI- SA dataset by setting the hyperparameters. The performance of the model is evaluated using various evaluation metrics of regression.

4.4 Validation and Model Evaluation

In this research, the Quantile regression technique is used for validating the temporal fusion transformer-based WQI prediction model. Quantile Regression is used to make probabilistic predictions at specific quantiles of the output distribution, instead of just a point estimate. Evaluation and validation of the model's performance are conducted utilizing a test set. The metrics such as mean absolute error, mean square error, root mean squared error and R2 score are used for evaluating the prediction model. Mean Absolute Error (MAE) is the average value of the absolute error, which reflects the true condition of the estimated value error more accurately. Mean Square Error (MSE) is the total and average of the square of the difference between the observed value and the forecasted value. Root Mean Square Error (RMSE) quantifies the difference between measured and real values. The R2 score value defines the accuracy of the prediction model, if the value is above 0.5 the model predicts efficiently

In this study, the effectiveness of the TFT-based WQI prediction models is determined by analyzing their error rate and R2 score value. The evaluation is conducted using a set of evaluation metrics, with 20% of the data being used as a test set.

5. Experiment and Results

In our prior study, the time series-based WQI-SA dataset composed of both physicochemical and seasonal parameters was utilized to train the WQI prediction models with deep neural architectures such as RNN, LSTM, and GRU. The performance of the WQI prediction models had been evaluated and observed good results for epoch size 200 as in Table 4. It was found that the LSTM-WQI prediction model had achieved a high R2 score value of 0.9 and less RMSE value of 0.3 when compared with RNN-WQI and GRU-WQI prediction models for epoch size 200.

Table 4. Prediction Results of Deep Learning Architecture

Model	Epoch	MAE	MSE	RMSE	R2 Score
RNN-WQI	200	0.25	0.25	0.5	0.84
LSTM-WQI		0.02	0.09	0.3	0.9
GRU-WQI		0.1	0.1936	0.44	0.88

In this research, the WQI prediction model is trained and tested with the same WQI-SA dataset. Various Python libraries are employed to conduct experiments and implement using temporal fusion transformer architecture. The dataset comprises 8124 tagged samples and represents 80% of the total instances of the WQI-SA dataset. The evaluation metrics such as mean absolute error, root mean squared error, and R2 score value are used to evaluate the prediction models. The test dataset containing 2009 instances is employed for testing. The performance of the WQI prediction model based on temporal fusion transformers is dependent on various factors such as the quality and quantity of the training data, the choice of model hyperparameters, and the ability of the model is to generalize to unseen data. Table 5 illustrates the general hyperparameter setting used in TFT training.

Table 5: Hyperparameters for fine-tuning

Hyperparameter	**Values**
Optimizer	Adam
Dense Layer	5 to 10
Epoch	20, 50, 100, 150, 200
Batch size	32/64

Hyperparameter	Values
Optimizer	Adam
Dense Layer	5 to 10
Dropout	0.2 or 0.3
Momentum	0.5 or 0.9
Learning Rate	0.1

The enhanced model employs the Adam optimization technique and divides the dataset into distinct sets for training, validation, and testing. This division allows for learning, hyperparameter tuning, and performance evaluation, respectively. The optimal model selection during training is facilitated through the utilization of early stopping. The exploration for improved TFT parameters encompasses a prediction time step of 30 steps, the encoder layer is set as 4, batch sizes are fixed to 64, state sizes from 32 to 256 and it is set to 64, learning rates from 0.0001 to 0.1, number of attention heads from 1 to 8, dropout rates from 0 to 0.4, Loss Function values from 0.1 to 0.9, Loss Function b values from 0.0001 to 0.5, and Loss Function g values from 0.1 to 0.5. The hyperparameters that boosted TFT of the two units are presented in Table 6.

Table 6: Special Hypermeter for TFT

Time steps	Encoders layers	Batch sizes	State size	Learning rates	Attention heads	Dropout rates	Loss Function a	Loss Function b	Loss Function g
30	4	64	64	0.01	4	0.20, 0.30	0.80	0.01	0.10

The results of the TFT architecture predictions have been analyzed for different dropout rates, including 0.2 and 0.3. It is observed that the dropout rate of 0.3 produced better prediction results. The mean absolute error for the TFT algorithm, with an epoch size of 200 and an optimizer of Adam, is 0.037 and the mean squared error is 0.01. Furthermore, the root mean squared error for the TFT-based WQI prediction model, with an epoch size of 200, was found to be 0.1 and the R2 score is determined to be 0.92.

The performance of the TFT architecture predictions is evaluated for various epoch sizes, including 20, 50, 100, 150, and 200. It is determined that an epoch size of 200 produced the best results. The mean absolute error for the TFT-based WQI prediction model, with an epoch size of 200 and an optimizer of Adam, is recorded as 0.037 and the mean squared error was 0.01. Additionally, the root mean squared error for the TFT-based WQI prediction model, with an epoch size of 200, is found to be 0.1, and the R2 score is determined to be 0.92.

Multiple experiments have been conducted using the WQI-SA dataset to develop WQI prediction models by varying the dropout rates (0.2 and 0.3) and the epoch sizes (20, 50, 100, 150, and 200). The results of these experiments, evaluated using standard metrics, are presented in Table 7a for the dropout rates and Table 7b for the epoch sizes.

Table 7a. Results of TFT-WQI Forecasts for Various Dropout Levels

Dataset	Dropout	MAE	MSE	RMSE	R2 Score
WQI-SA	0.3	0.037	0.01	0.1	0.92
	0.2	0.12	0.15	0.387	0.89

Table 7b. Outcomes of TFT-WQI Predictions Across Different Epochs

Dataset	**Epochs**	**MAE**	**MSE**	**RMSE**	**R2 Score**
WQI-SA	200	**0.037**	**0.01**	**0.1**	**0.92**
	150	0.083	0.136	0.368	0.9
	100	0.12	0.15	0.387	0.89
	50	0.27	0.2	0.44	0.88
	20	0.34	0.23	0.479	0.86

Fig.4 visualizes the actual time series data and the predicted time series through the use of quantile bands. The bands are determined by the values qL1, qU1, qL2, qU2, qL3, and qU3, which define the lower and upper bounds of each quantile. The actual time series is represented by a line labelled actual, while the predicted time series is represented by three distinct quantile bands, each defined by qL and qU values and labelled with the corresponding string. The expected value of the predicted time series, computed as the median as central quantile = mean, is plotted and labelled as expected.

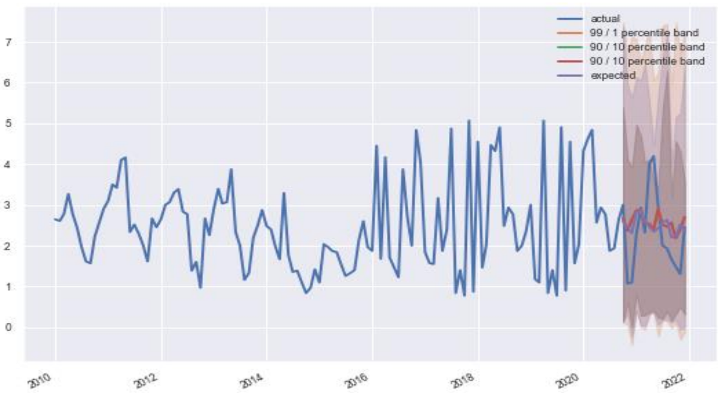

Fig.4. Visualizing Actual and Predicted Time Series Data with Quantile Bands

Comparative Analysis

This study conducts a comparative analysis of various water quality index (WQI) prediction models. The performance of the WQI prediction model based on temporal fusion transformer (TFT) architectures is compared with the LSTM, GRU, and RNN architectures using metrics such as MAE, MSE, RMSE, and R2 score. The TFT model is implemented for 20, 50, 100, 150, and 200 epochs with hyperparameters set into the model. The LSTM-WQI, GRU-WQI, and RNN-WQI models are also executed for the same number of epochs using relevant hyperparameters. The comparative results of the TFT-WQI prediction model are compared with deep learning architectures and depicted in Table 8 and the performance analysis is illustrated in Fig.5.

Table 8. Comparative Performance Result of WQI Prediction Models

Model	Dropout	Epoch	MAE	MSE	RMSE	R2 Score
TFT-WQI	**0.3**	**200**	**0.037**	**0.01**	**0.1**	**0.92**
RNN -WQI			0.23	0.25	0.5	0.84
LSTM-WQI			0.02	0.09	0.3	0.9
GRU-WQI			0.1	0.2	0.44	0.88

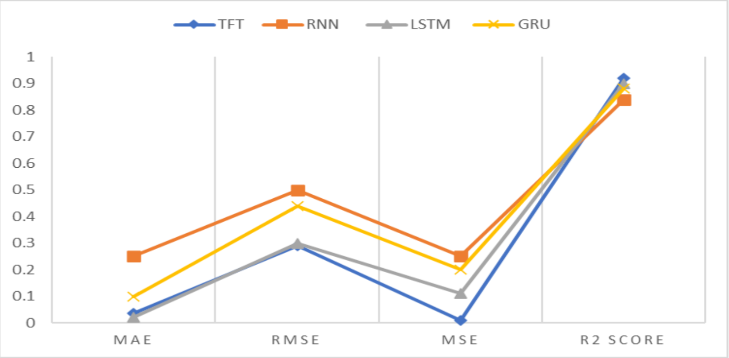

Fig.5. Performance Analysis of WQI Prediction Models

The mean absolute error is observed as 0.037 for TFT based WQI prediction model, whereas 0.25 for RNN, 0.02 for LSTM and 0.1 for GRU-based prediction models. High MAE is observed for GRU based prediction model and less MAE is obtained for the TFT-WQI prediction model and the results are illustrated in Fig.6a.

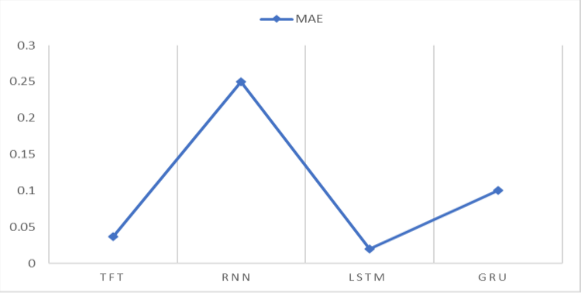

Fig. 6a. MAE of WQI Prediction Models

The RMSE for the WQI prediction model based on TFT architectures is observed to be 0.1, the other deep learning architectures such as RNN-WQI obtained 0.5, LSTM-WQI is 0.3 and GRU-WQI is 0.44 and is depicted in Fig.6b. The high RMSE is observed for the RNN-based WQI prediction model and less error is obtained when employing TFT-WQI prediction model for the given dataset.

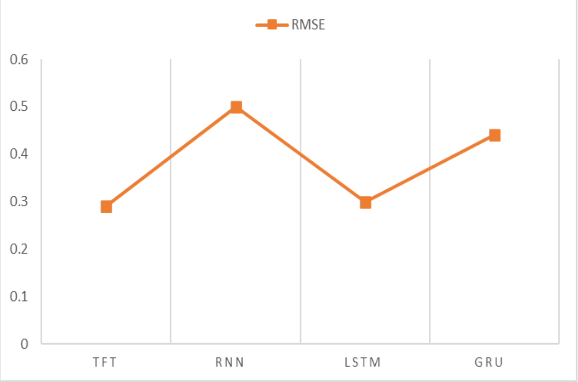

Fig 6b. RMSE of WQI Prediction Models

The R2 score value for TFT based WQI prediction model is observed as 0.92 and the outperforming model in river water quality forecasting. The deep learning algorithms are compared with the TFT WQI prediction model to find the best efficient model. The LSTM-WQI approach is found 0.9, RNN-WQI got 0.84 and the GRU-WQI model is observed at 0.88 while training with the WQI-SA dataset and the results are illustrated in Fig.6c.

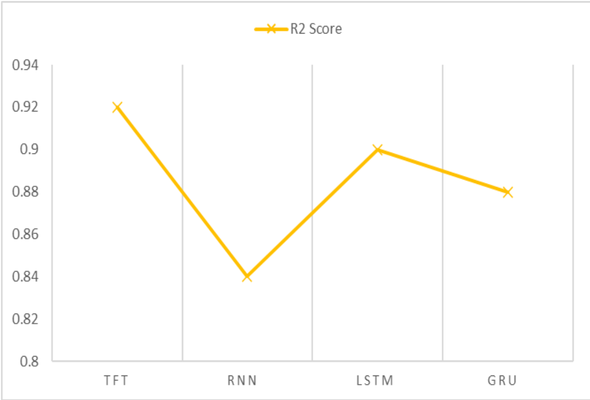

Fig 6c. R2 Score of WQI Prediction Models

Findings

The results of the present study conclusively demonstrate the efficacy of the Temporal Fusion Transformer architecture in the development of prediction models for time series data, such as the prediction of water quality index. The WQI-prediction model built with TFT is found that the TFT-WQI prediction model handled time series-based water quality data efficiently with a high R2 score value and less error rate. The investigations carried out in this research provide strong evidence of the usefulness of the TFT architecture in addressing the challenges inherent in the prediction of WQI. The Temporal Fusion Transformer-based WQI prediction model has been found to have several key advantages over traditional models, including improved accuracy, better handling of temporal dependencies, better representation of complex relationships, and improved generalization performance. The incorporation of specific hyperparameters has led to improvements in the TFT architecture, making it stand apart from other models. The temporal fusion approach effectively captures the temporal relationships between different water quality parameters, while the Transformer architecture is well-suited for the handling of river water quality data. The resulting model provides a clear and concise understanding of the relationships between different water quality parameters and is better equipped to handle complex and diverse data sets. The model demonstrates strong generalization performance and allows to effectively predict the water quality index in real-world scenarios.

6. Conclusion

This research highlights the effectiveness of the Temporal Fusion Transformer approach in predicting the water quality index. The study utilized seasonal data obtained from a visual crossing site between the years 2016 and 2020, integrated it with the physiochemical parameters of the Bhavani River water and resulted in the construction of a novel time series dataset. The Temporal Fusion Transformer (TFT) approach was utilized to design and develop a river water quality forecasting model. The performance of the TFT model was evaluated and compared with LSTM-WQI, RNN-WQI and GRU-WQI prediction models and it was found that the TFT-WQI prediction model handled time series-based water quality data efficiently. The research determined that the utilization of Temporal Fusion Transformer architecture represents a meaningful advancement in water quality prediction by demonstrating the efficacy in forecasting the Water Quality Index. Furthermore, the research developed a generalized TFT model that can be applied to predict the water quality of any river and can also serve as a pre-trained model for transfer learning, which is a significant step forward in the field of water quality prediction.

Reference

[1] Chang, H., Zhang, Y., & Chen, L. (2003). Gray forecast of Diesel engine performance based on wear. Applied Thermal Engineering, 23(17), 2285–2292.

[2] Ait Maatallah, O., Achuthan, A., Janoyan, K., & Marzocca, P. (2015). Recursive wind speed forecasting based on Hammerstein Auto-Regressive model. Applied Energy, 145, 191–197.

[3] Chiang, C.J.; Yang, J.L.; Cheng, W.C. (2013) Temperature and state-of-charge estimation in ultracapacitors based on extended Kalman filter. J. Power Sour. 234, 234–243.

[4] Lee, S.; Kim, C.K.; Kim, D.(2020) Monitoring Volatility Change for Time Series Based on Support Vector Regression. Entropy, 22, 1312.

[5] Gumus, M.; Kiran, M.S. Crude oil price forecasting using XGBoost. In Proceedings of the 2017 International Conference on Computer Science and Engineering (UBMK), Antalya, Turkey, pp. 1100–1103.

[6] Hotait, H.; Chiementin, X.; Rasolofondraibe, L.(2021) Intelligent Online Monitoring of Rolling Bearing: Diagnosis and Prognosis. Entropy, 23, 791.

[7] Dudek, G. Short-term load forecasting using random forests. (2014) In Proceedings of the 7th IEEE International Conference Intelligent Systems IS'2014 (Advances in Intelligent Systems and Computing), Warsaw, Poland

[8] Muzaffar, S.; Afshari, (2019), A. Short-term load forecasts using LSTM networks. Energy Procedia, 158, 2922–2927.

[9] Jiang, Q.; Tang, C.; Chen, C.; Wang, X.; Huang, Q. Stock price forecast based on LSTM neural network.(2018) In Proceedings of the Twelfth International Conference on Management Science and Engineering Management, Melbourne, Australia; pp. 393–408.

[10] Hewage, P.; Behera, A.; Trovati, M.; Pereira, E.; Ghahremani, M.; Palmieri, F.; Liu, Y. (2020) Temporal convolutional neural (TCN) network for an effective weather forecasting using time-series data from the local weather station. Soft Comput, 16453–16482.

[11] Lim, B., Arik, S. O., Loe , N. & Pster, T.(2021) Temporal fusion transformers for interpretable multi-horizon time series forecasting. International Journal of Forecasting.

[12] Li, Z. et al. (2018) Water Quality Prediction Model Combining Sparse Auto-encoder and LSTM Network. 51 (17), 831.

[13] Srivastava, A., & Cano, A. (2022). Analysis and forecasting of rivers pH level using deep learning. Progress in Artificial Intelligence, 11(2), 181–191.

[14] Yang, Y. et al. (2021), A study on water quality prediction by a hybrid CNN-LSTM model with attention mechanism. Environmental Science and Pollution Research.

[15] J. P. Nair and M. S. Vijaya, "Predictive Models for River Water Quality using Machine Learning and Big Data Techniques - A Survey,"International Conference on Artificial Intelligence and Smart Systems (ICAIS), 2021, pp. 1747-1753.

[16] Aldhyani THH, Al-Yaari M, Alkahtani H, Maashi M. Water Quality Prediction Using Artificial Intelligence Algorithms. Appl Bionics Biomech. 2020

[17] Nair, Jitha P., and M. S. Vijaya. (2022) River Water Quality Prediction and Index Classification Using Machine Learning'. Journal of Physics: Conference Series, vol. 2325, no, p. 012011.

[18] Nair, J.P., Vijaya, M.S. (2022). Exploratory Data Analysis of Bhavani River Water Quality Index Data. In: Kumar, S., Hiranwal, S., Purohit, S.D., Prasad, M. (eds) Proceedings of International Conference on Communication and Computational Technologies. Algorithms for Intelligent Systems. Springer, Singapore.

[19] Heddam, S., 2014. Generalized regression neural network-based approach for modelling hourly dissolved oxygen concentration in the Upper Klamath River, Oregon, USA. Environmental Technology (United Kingdom) 35, 1650–1657. 10.1080/ 09593330.2013.878396.

[20] Basant, N., Gupta, S., Malik, A., Singh, K.P., 2010. Linear and nonlinear modelling for simultaneous prediction of dissolved oxygen and biochemical oxygen demand of the surface water case study Chemometr.Intellig.Lab.Syst.104,172–180.

[21] Wang, Y., Zhou, J., Chen, K., Wang, Y., & Liu, L. (2017). Water quality prediction method based on LSTM neural network. In 2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE) (pp. 1-5). IEEE.

[22] Santhana Lakshmi, V., Vijaya, M.S. (2022). A Study on Machine Learning-Based Approaches for PM2.5 Prediction. In: Karrupusamy, P., Balas, V.E., Shi, Y. (eds) Sustainable Communication Networks and Application. Lecture Notes on Data Engineering and Communications Technologies, vol 93. Springer, Singapore.

[23] Heddam, S., Kisi, O., 2018. Modelling daily dissolved oxygen concentration using least square support vector machine, multivariate adaptive regression splines and M5 model tree. J. Hydrol. 559, 499–509.

[24] Nair, Jitha P., and M S Vijaya. (2022) 'Analysing And Modelling Dissolved Oxygen Concentration Using Deep Learning Architectures'. International Journal of Mechanical Engineering, vol. 7, pp. 12–22.

[25] Li L, Jiang P, Xu H, Lin G, Guo D, Wu H (2019) Water quality prediction based on recurrent neural network and improved evidence theory: a case study of Qiantang River, China. Environ Sci Pollut Res 26(19): 19879–19896

[26] Fatima Abbas, Deep Learning Approaches for Medical Image Analysis and Diagnosis , Machine Learning Applications Conference Proceedings, Vol 3 2023.

[27] Paul Garcia, Ian Martin, Laura López, Sigurðsson Ólafur, Matti Virtanen. Enhancing Student Engagement through Machine Learning: A Review.

Kuwait Journal of Machine Learning, 2(1). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/163

[28] Sai Pandraju, T. K., Samal, S., Saravanakumar, R., Yaseen, S. M., Nandal, R., & Dhabliya, D. (2022). Advanced metering infrastructure for low voltage distribution system in smart grid based monitoring applications. Sustainable Computing: Informatics and Systems, 35 doi:10.1016/j.suscom.2022.100691