6. TEMPORAL FUSION TRANSFORMER FOR WQI PREDICTION MODEL

In this research, deep learning models have emerged as a promising and viable alternative, utilizing diverse data sources, including physiochemical and seasonal data, for building WQI prediction models. Temporal Fusion Transformer is a powerful deep learning architecture that efficiently handles time-series data by incorporating temporal and cross-dimensional attention mechanisms. This architecture has shown promising results in various applications, including predicting complex and dynamic patterns in time-series data. In this context, an enhanced WQI prediction model using multi-horizon forecasting with a temporal fusion transformer is proposed in this work. This chapter demonstrates the application of TFT for building WQI prediction model by training the pooled features of Bhavani River data.

WQI PREDICTION MODEL USING TEMPORAL FUSION TRANSFORMER

In this study, a new method for modelling and forecasting river WQI using temporal fusion transformers is proposed. TFTs are a type of deep learning model used to capture the temporal dependencies between time series data. The objective of this work is to build an enhanced WQI prediction model using multi-horizon forecasting with a temporal fusion transformer. For training TFT, the time series river water quality data is used and various hyperparameters are appropriately defined while creating the model. The task of WQI prediction is approached as a regression problem, and the regression model is constructed by leveraging insights gained from the training data using TFT architectures.

Methodology

Temporal Fusion Transformer is an advanced architecture that combines the power of transformer networks with time-series forecasting capabilities. It effectively models temporal dependencies in sequential data, making it suitable for time-series forecasting tasks. TFT employs self-attention mechanisms to capture long-range dependencies and learns to fuse information from multiple temporal resolutions, enabling it to handle irregular time intervals in the input data. The proposed methodology consists of the key components such as 1. data collection, 2. EDA and data pre-processing 3. construction of the WQI prediction model, and 4. model evaluation. The framework of the WQI prediction model built based on TFT architecture is illustrated in Fig.6.1.

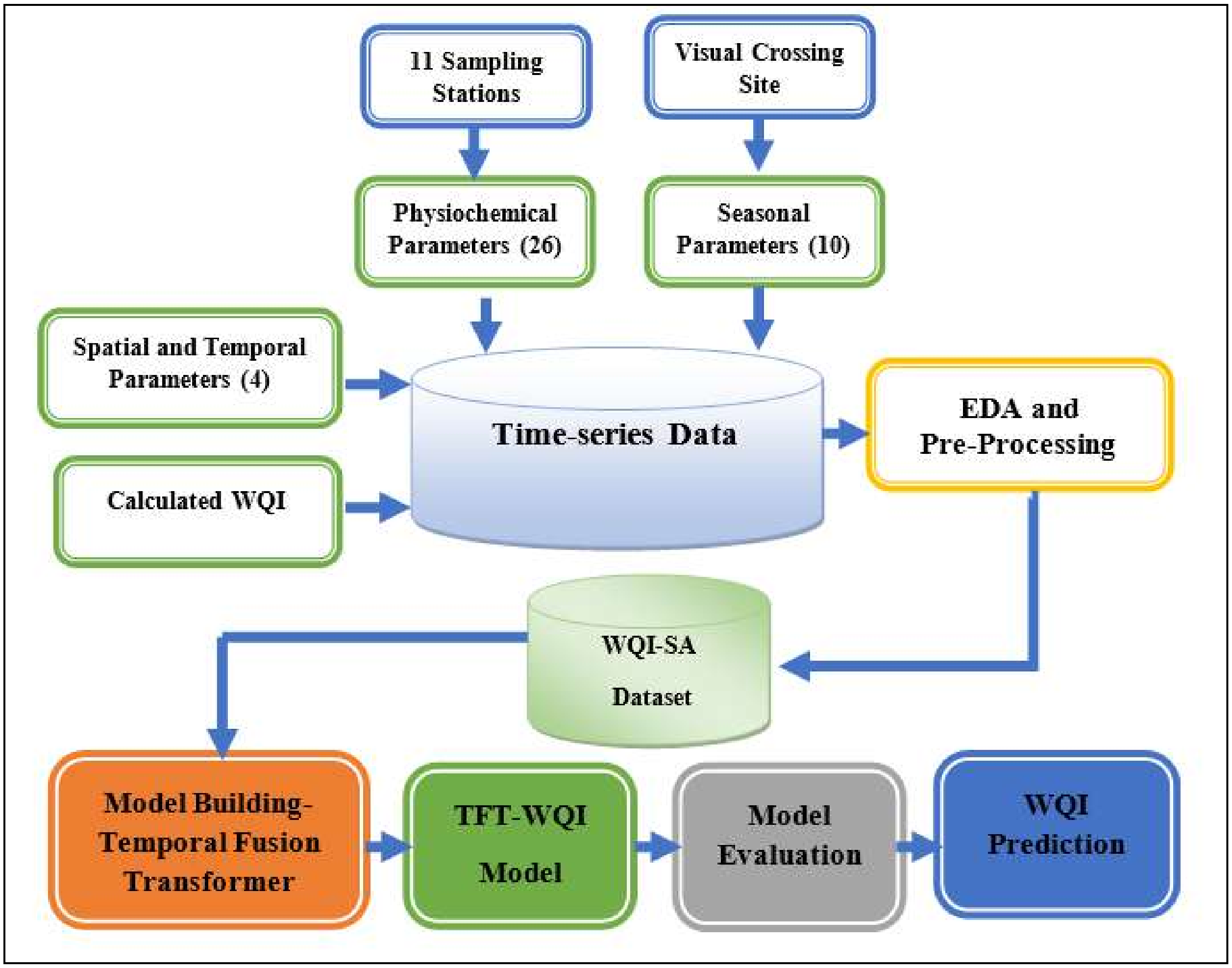

Fig. 6.1 Framework of the WQI Prediction Model Based on TFT Architecture

Data Collection and Dataset Preparation

A total of 10,560 water samples were gathered from 11 monitoring stations situated along the Bhavani River. The data collection period spanned from January 1st, 2016 to December 31st, 2020. A time series data containing 26 physiochemical parameters, 10 seasonal parameters, longitude, latitude, station ID, date and calculated WQI has been developed. EDA is conducted on the time series data to gain insights into its characteristics and to analyse the significance of each parameter in determining the water quality index. The distribution of parameter values is studied and comprehended using a range of statistical techniques, including heatmap analysis, boxplot analysis, pair plot analysis, and histogram analysis. Based on the findings of EDA, certain preprocessing requirements are identified and subsequently implemented. The Select K best feature selection method is applied to retain only the most relevant features. The application of the

Select K Best feature selection method yielded significant improvements in the river water quality dataset, which finally comprising 10,560 tagged instances and 38 attributes, and this dataset is called as the WQI-SA dataset as mentioned in Table XII of Chapter 3.

Model Building

The task of predicting the water quality index is approached as a regression problem and is tackled using TFT architectures. The WQI prediction model are developed using temporal fusion transformer architecture by training the 80% tagged instances of WQI-SA dataset. 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 feature representations of encoders. 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.

During training, TFTs utilize temporal and cross-dimensional attention mechanisms to effectively capture complex temporal patterns and relationships within the water quality data. The model processes the sequential instances and attributes of the dataset, learning to weigh their importance dynamically over different time steps. The model is trained using the specified number of epochs, and the learning rate is set to the specified value.

The various hyperparameters such as the number of hidden layers, the number of neurons in each layer, and the learning rate are set while training. 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. These special hyperparameters are important for TFT to perform well on sequence data and significantly impact the performance of the model. Thus, an enhanced WQI prediction model is built by learning the trends in water quality parameters from the WQI-SA dataset with a temporal fusion transformer by properly setting and fine tuning the hyperparameters and the model is referred to as the TFT-WQI-SA model. The predictive performance of WQI model is evaluated using the metrics such as MAE, MSE, RMSE, and R2 score.

Experiments and Results

The experiments are conducted by implementing the TFT deep learning architecture and implementing it using Python libraries under TensorFlow and Keras. The training dataset comprised 8124 labelled instances of the WQI-SA dataset. The evaluation of the prediction performance of the model is undertaken using various metrics, such as MAE, MSE, RMSE and R2 score values, utilizing a separate test dataset containing 2009 instances.

The performance results of the WQI prediction model built using temporal fusion transformers depends on a variety of factors depends on the choice of model hyperparameters as tabulated in Table XXVIII. The setting of special hyperparameters of TFT 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 0.2 to 0.3 as tabulated in Table XXIX.

Table XXVIII. Normal Hyperparameters Setting for Training TFT

Hyperparameter	Values	Hyperparameter	Values
Optimizer	Adam	Dropout	0.2, 0.3
Dense Layer	5 to 10	Momentum	0.5 or 0.9
Epoch	20, 50, 100, 150, 200	Learning rate	0.1
Batch size	32/64	Activation function	Relu

Table XXIX: Setting of Special Hyperparameters for TFT

Time steps	30
Encoders Layers	4
Batch Size	64
State size	64
Learning rates	0.01
Attention heads	4
Dropout rate	0.20, 0.30
Loss Function a	0.8
Loss Function b	0.01
Loss Function g	0.1

The results of the TFT-based WQI prediction model (TFT-WQI-SA model) are experimented with various epochs such as from 20 to 500 where various metrics are measured at different epochs. For the WQI-SA dataset, at epoch 500, the TFT-WQI-SA model achieves an MAE value of 0.122, which represents the average absolute difference between the predicted and actual values. The MSE is calculated as 0.167, signifying the average of squared differences. The RMSE, derived as the square root of the MSE, is reported as 0.4087. The R2 Score, is noted at 0.941, suggesting a high level of prediction performance. At 200 epochs, the MAE experiences a marginal increase to 0.137, indicating a slightly larger average absolute difference in predictions. Simultaneously, the MSE rises to 0.183, signifying a higher average of squared differences. The RMSE value expands to 0.4278, while the R2 Score decreases slightly to 0.928.

Continuing to 150 epochs, the MAE exhibits a more pronounced increase to 0.183, and the MSE follows suit, expanding to 0.197. The RMSE at this stage reaches 0.4438. The R2 Score experiences a further reduction to 0.913. Upon reaching 100 epochs, the MAE increases again to 0.214, and the MSE follows suit, expanding to 0.218. The RMSE value rises to 0.4669, and the R2 Score decreases further to 0.89. As the number of epochs decreases to 50, the MAE exhibits a larger increase to 0.227, accompanied by a higher MSE of 0.276. The RMSE now stands at 0.5254, and the R2 Score experiences a more substantial decline to 0.887. Finally, with only 20 epochs, the MAE maintains a higher value of 0.248, and the MSE increases to 0.293. The RMSE value stands at 0.5413, while the R2 Score decreases further to 0.87. The performance analysis of the TFT-WQI-SA model on the WQI-SA dataset at different epochs is tabulated in Table XXX.

Table XXX. Performance Results of TFT-WQI-SA Model for Different Epochs

Dataset	Epochs	MAE	MSE	RMSE	R2 Score
WQI-SA	**500**	0.122	0.167	0.4087	**0.941**
	200	0.137	0.183	0.4278	0.928
	150	0.183	0.197	0.4438	0.913
	100	0.214	0.218	0.4669	0.89
	50	0.227	0.276	0.5254	0.887
	20	0.248	0.293	0.5413	0.87

The results of the TFT-based WQI prediction model (TFT-WQI-SA model) have experimented with different drop out such as 0.2 and 0.3 where various metrics are measured. For a dropout rate of 0.3, the TFT-WQI-SA model achieves an MAE of 0.122, which represents the average absolute difference between the predicted and actual values. The MSE is calculated as 0.167, representing the average of squared differences. The RMSE is 0.4087, which is the square root of the MSE. The R2 score, measuring the goodness of fit, is 0.941, indicating a very high level of prediction accuracy. Decreasing the dropout rate to 0.2 leads to a higher MAE of 0.183, suggesting a larger difference between the predicted and actual values. The MSE increases to 0.197, and the RMSE becomes 0.4438. The R2 score decreases to 0.913, indicating a slightly lower level of prediction accuracy compared to the dropout rate of 0.3. The performance analysis of dropout is tabulated in Table XXXI.

Table XXXI. Performance of TFT-WQI-SA Model for Different Dropout Rates

Dataset	Dropout	MAE	MSE	RMSE	R2 Score
WQI-SA	0.3	0.122	0.167	0.4087	0.941
	0.2	0.183	0.197	0.4438	0.913

The actual time series data and the predicted time series through the use of quantile bands are depicted in Fig.6.2. 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 the central quantile mean, is plotted and labelled as expected.

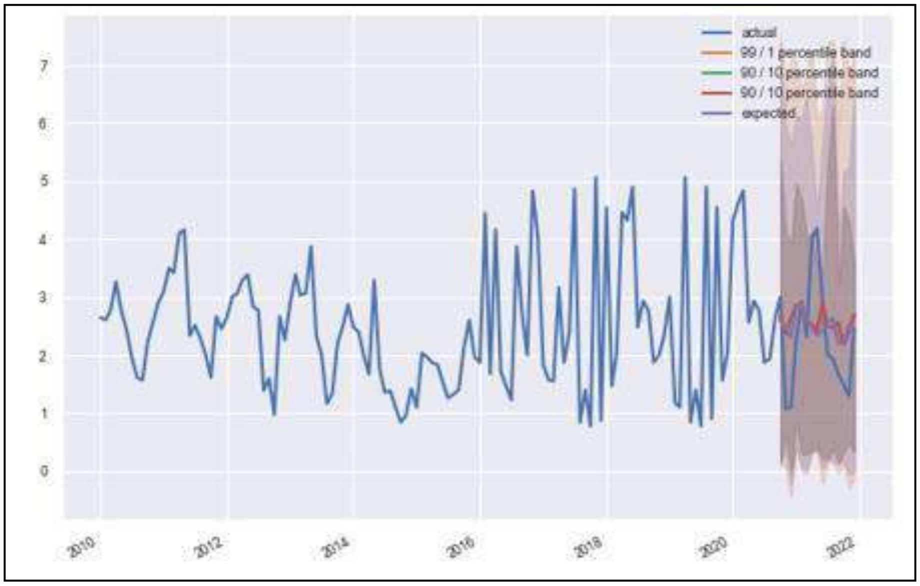

Fig.6.2. Visualization of Actual and Predicted Values with Quantile Bands

Comparative Analysis

The results of the TFT-WQI-SA prediction model are compared with the performance results of the prediction models such as RNN-WQI-SA, LSTM-WQI-SA and GRU-WQI-SA described in Chapter 5. The hyperparameters for the TFT-WQI-SA prediction model are set into the model, which is then implemented for 20, 50, 100, 150, 200 and 500 epochs. The RNN-WQI-SA, LSTM-WQI-SA, and GRU-WQI-SA prediction model**s** are also executed for the same number of epochs, using relevant hyperparameters.

The MAE is observed as 0.122 for TFT based WQI prediction model, whereas 0.428 for RNN, 0.298 for LSTM and 0.39 for GRU-based prediction models. High MAE is observed for GRU based prediction model and less MAE is obtained for the TFT-WQI-SA prediction model. The MSE for the WQI prediction model based on TFT-WQI-SA is observed to be 0.167, the other deep learning architectures such as RNN-WQI-SA obtained 0.384, LSTM-WQI-SA is 0.2084 and GRU-WQI-SA is 0.2149. The high MSE is observed for the RNN-WQI-SA prediction model and less error is obtained for the TFT-WQI-SA prediction model for the given dataset.

The RMSE for the WQI prediction model based on TFT-WQI-SA is observed to be 0.4087, the other deep learning architectures such as RNN-WQI-SA obtained 0.6197, LSTM-WQI-SA is 0.4565 and GRU-WQI-SA is 0.4636. The high RMSE is observed for the RNN-WQI-SA prediction model and less error is obtained by the TFT-WQI-SA prediction model for the given dataset. The R2 score value for the TFT-WQI-SA prediction model is observed as 0.941 and the outperforming model in river water quality forecasting. The LSTM-WQI-SA approach is found 0.82, RNN-WQI-SA got 0.856 and the GRU-WQI-SA model is observed at 0.839 while training with the WQI-SA dataset.

The comparative performance results of the TFT-WQI-SA prediction model and RNN-WQI-SA, LSTM-WQI-SA and GRU-WQI-SA models are depicted in Table XXXII and the performance analysis is illustrated in Fig.6.3.

Table XXXII. Performance of Different WQI Models Based on WQI-SA Dataset

Model	Dropout	Epoch	MAE	MSE	RMSE	R2 Score
TFT-WQI-SA	**0.3**	**500**	**0.122**	**0.167**	**0.4087**	**0.941**
RNN-WQI-SA			0.428	0.384	0.6197	0.82
LSTM-WQI-SA			0.298	0.2084	0.4565	0.856
GRU-WQI-SA			0.39	0.2149	0.4636	0.839

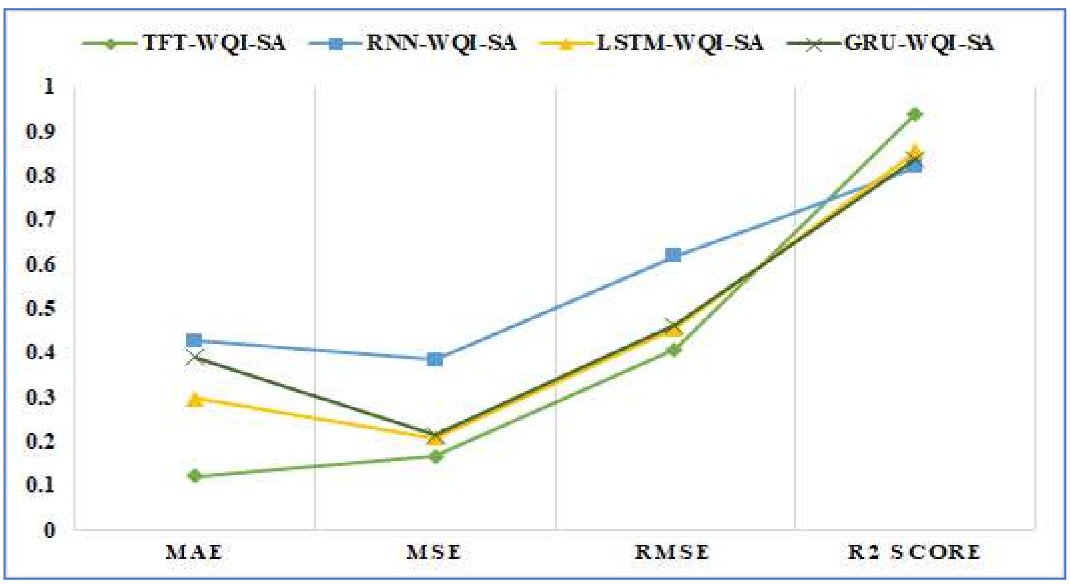

Fig.6.3. Performance Analysis of the Various WQI Prediction Models Based on WQI-SA Dataset

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 found that the model identified and handled time series patterns of water quality data efficiently with a high R2 score value and less error rate. 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 TFT-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 WQI prediction, making it stand apart from other models. The model demonstrates strong generalization performance and can effectively predict the water quality index in real-world scenarios.

SUMMARY

This chapter presents the methodology of building an enhanced WQI prediction model built with specially designed time series architecture TFT. The implementation and model training using TFT architecture for predicting the WQI has been described in detail with experimental results. The performance of TFT-WQI-SA prediction models is compared with the models built in Chapter 5 and the comparative performance results have been reported. To further generalize the method of building the WQI prediction model, a transfer learning approach is proposed in this research and the construction of the WQI prediction model with homogenous transfer learning will be explained in the next chapter.

Remarks

The paper titled 'Temporal Fusion Transformer: A Deep Learning Approach for Modelling and Forecasting River Water Quality Index' has been published in the International Journal of Intelligent Systems and Applications in Engineering, Vol.11(10s), 2023.

(Scopus Indexed)