7. HOMOGENEOUS TRANSFER LEARNING FOR WQI PREDICTION MODELS

The main drawback of machine learning lies in its dependency on limited datasets for training, leading to reduced accuracy. More extensive and diverse data is required to effectively train models and enhance their performance thereby overcome the limitation of limited data. In real-time applications, obtaining large and diverse datasets for machine learning is challenging due to various reasons, such as data availability, privacy concerns, and location-dependency. Certain locations may have limited data availability, making it difficult to gather enough relevant information to train accurate models. Even though deep learning architectures excel at handling big data and complex tasks, their performance may suffer when faced with a scarcity of data, leading to suboptimal accuracy. The effectiveness of these models heavily relies on having a substantial and diverse dataset for training to achieve desirable results.

In this research two rivers such as Bhavani River and Bharathapuzha river have been taken for study, such that a large number of samples are available for Bhavani River as compared to the Bharathapuzha river which has a smaller number of instances. A new approach called transfer learning has been evolved to achieve good performance with best recognition rate, which uses the knowledge acquired from the pretrained model along with the given training samples to train the network and thereby constructs an improved model. The transfer learning approach facilitates the development of a more generalized model, particularly beneficial in scenarios involving limited data. Transfer learning offers a valuable advantage in optimizing model development by leveraging pre-trained knowledge from one domain to enhance performance in a different, related domain. This approach significantly reduces the need for extensive data and computational resources while achieving better generalization and faster convergence. Hence in this work, a transfer learning approach is applied for training and building the WQI prediction model to achieve better performance. By leveraging the knowledge from the WQI model built with Bhavani River dataset, the model effectively adapts and improve the performance when applied to the Bharathapuzha river dataset, which has limited samples.

The chapter begins with the description of building the base models for Bharathapuzha River by employing the RNN variants and the TFT. The application of transfer learning with LSTM based pre-trained model and training the architectures RNN, LSTM and GRU with

Bharathapuzha river for building WQI prediction models is presented in the 2nd section. The application of transfer learning with TFT based pre-trained model and training the architectures RNN, LSTM, GRU, and TFT with Bharathapuzha river for building WQI prediction models is presented in the 3rd section.

7.1. WQI PREDICTION MODEL USING BHARATHAPUZHA RIVER DATA

In this work, WQI prediction model for the Bharathapuzha River is developed by learning time series data containing physiochemical and seasonal parameters using recurrent neural network architectures, RNN, LSTM, and GRU. Also, the WQI model is built by training TFT architecture. The methodology described in Chapter 4 to Chapter 6 is adopted here also to implement the regression task for building the WQI prediction model but the time series data of Bharathapuzha River, which is limited in size, is used for modelling and training the neural networks.

Methodology

RNN, LSTM, GRU, and TFT are highly valuable architectures in sequence and time series data analysis and training due to their ability to effectively capture temporal dependencies and patterns. 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.7.1.

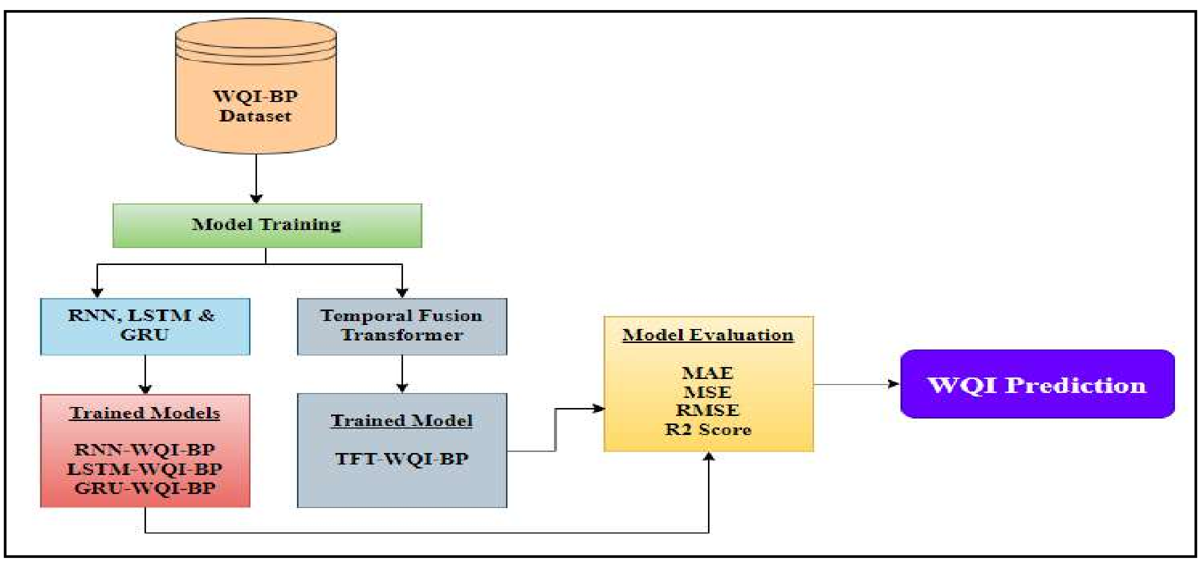

Fig.7.1. Framework of WQI Prediction Model (base model) for Bharathapuzha River

Bhrathapuzha River Dataset (WQI-BP)

The river water quality samples comprise 26 physicochemical parameters that have been collected from three monitoring stations along the Bharathapuzha River, and 10 seasonal characteristics recorded at visual crossing sites. The water quality index value for each sample is calculated and assigned to the corresponding samples as a target variable. A time series data with 2190 and 41 attributes including 26 physiochemical parameters, 10 seasonal parameters, longitude, latitude, station ID, date and calculated WQI, has been created.

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 2190 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 deep learning architectures such as RNN, LSTM and GRU excel in characterizing and modelling data inputs through their interconnected nodes with multiple hidden layers. The model building process involved constructing each network with the integration of the WQI-BP data. The training of these models has been performed using the pre-processed data samples of the Bharathapuzha River. The WQI prediction models have been trained and developed by leveraging the temporal dependencies and contextual patterns present in the WQI-BP dataset, enabling them to effectively capture the complex relationships between water quality parameters. Hyperparameters such as epochs, dropout, learning rate, momentum and etc. are defined and fine-tuning is done and optimize models. Three distinct WQI prediction models have been developed and named as RNN-WQI-BP, LSTM-WQI-BP, and GRU-WQI-BP.

Again, the Temporal Fusion Transformer, a powerful framework designed to handle time series data with complex temporal dependencies has been trained with instances of the WQI-BP dataset. TFT combines transformer blocks to capture long-term dependencies in time series data.

The input layer processes the WQI-BP time series data, converting it into numerical representations. Encoder layers employ self-attention mechanisms to generate feature representations for each time step, incorporating information from past and future steps. The TFT architecture effectively captures and leverages the temporal patterns and dependencies in the water quality parameters while training. Hyperparameters such as epochs, dropout, learning rate, momentum and etc and special hyperparameters such as encoder layers, batch size, and attention layers are properly configured and an efficient WQI prediction model is built and named as TFT-WQI-BP model.

In the process of building the above WQI prediction models RNN-WQI-BP, LSTM-WQI-BP, GRU-WQI-BP and TFT-WQI-BP, 80% of the prepared WQI-BP dataset is utilized as input to train the networks. The effectiveness of the models RNN-WQI-BP, LSTM-WQI-BP, GRU-WQI-BP and TFT-WQI-BP in forecasting the water quality index is assessed through metrics such as Mean Absolute Error, Mean Squared Error, Root Mean Squared Error, and the R2 score value, with 20% of the instances of WQI-BP dataset as test samples.

Experiment and Results

The experiments have been carried out by training the Bharathapuzha River water dataset using deep learning algorithms such as RNN, LSTM, and GRU. Also experimented with TFT architecture and implemented using Python libraries. The training dataset with 1752 instances of the river water dataset covering has been used for training the networks. The test set with 438 instances has been used for testing the performance of the prediction models and evaluated for its efficiency in forecasting the water quality using the MAE, MSE, RMSE and R2 score values. The deep learning structures RNN, LSTM, and GRU are configured with hyperparameters as mentioned in Table XXVII of Chapter 6, such as dense layer values from 5 to 10 units, optimizer as adam optimizer. The epoch sizes were listed as 20, 50, 100, 150, 200 and 500. The activation functions are defined as relu, and the momentum is set between 0.5 and 0.9. The dropout unit is 0.2 or 0.3, the learning rate is 0.1, and the batch size is set at either 32 or 64, the setting of the hyperparameter for the building model. During the experiments, it is found that the results stabilized with momentum as 0.8, epoch sizes 500, drop out 0.3 and with relu activation function and achieved better results.

The performance of the RNN-based WQI prediction model (RNN-WQI-BP model) is assessed by conducting experiments with a range of epochs, varying from 20 to 500. At 20 epochs, the model exhibits an MAE of 0.647, MSE of 0.662, RMSE of 0.8136, and R2 score of 0.584. With the progression to 50 epochs, a noticeable reduction is observed in MAE of 0.638, MSE of 0.642, RMSE of 0.8012, and a slight improvement in R2 score of 0.589. Continuing the training process to 100 epochs results in further refinement of predictions, as indicated by the diminishing MAE of 0.631, MSE of 0.628, and RMSE of 0.7925 values, accompanied by a modest increase in R2 score of 0.593.

As the training continues, particularly at 150 epochs, the predictive accuracy is enhanced, demonstrated by the decreasing MAE of 0.627, MSE of 0.612, and RMSE of 0.7823 values, with a corresponding improvement in R2 score of 0.606. The trend of refinement persists at 200 epochs, where the model attains lower MAE of 0.614, MSE of 0.584, and RMSE of 0.7642 values, contributing to a higher R2 score of 0.627. At the maximum epoch value of 500, the model predictive prowess becomes evident, as demonstrated by the notably reduced MAE of 0.598, MSE of 0.523, RMSE of 0.7232, and a favourable R2 score of 0.642. The performance evaluation of the WQI prediction model based on the WQI-BP dataset with traditional RNN-WQI-BP is shown in Table XXXIII.

Table XXXIII. Prediction Results of RNN-WQI-BP Model for Various Epochs

Dataset	Epochs	MAE	MSE	RMSE	R2 Score
WQI-BP	20	0.647	0.662	0.8136	0.584
	50	0.638	0.642	0.8012	0.589
	100	0.631	0.628	0.7925	0.593
	150	0.627	0.612	0.7823	0.606
	200	0.614	0.584	0.7642	0.627
	500	0.598	0.523	0.7232	0.642

The performance of the LSTM-based WQI prediction model (LSTM-WQI-BP model) is assessed by conducting experiments with a range of epochs, varying from 20 to 500. At 20 epochs, the model exhibits a MAE of 0.591, MSE of 0.613, RMSE of 0.7829, and an R2 score of 0.592. As the training process advances to 50 epochs, a gradual enhancement is observed in the model performance, leading to decreased MAE of 0.585, MSE of 0.596, and RMSE of 0.7720, accompanied by a marginally improved R2 score of 0.604. Continuing the training to 100 epochs demonstrates further refinement in predictive accuracy, highlighted by a lower MAE of 0.563, MSE of 0.583, and RMSE of 0.7635, while showcasing a notable improvement in the R2 score of 0.627. At 150 epochs, the model continues to progress, with reduced MAE of 0.552, MSE of 0.547, and RMSE of 0.7396, leading to an elevated R2 score of 0.636. As the training iterations reach 200 epochs, reflecting decreased MAE of 0.543, MSE of 0.535, and RMSE of 0.7314, resulting in a higher R2 score of 0.657. Finally, at 500 epochs, the model reaches a noteworthy level of predictive accuracy, evidenced by lower MAE of 0.524, MSE of 0.513, and RMSE of 0.7162, culminating in an impressive R2 score of 0.684. The performance evaluation of the LSTM model is shown in Table XXXIV.

Table XXXIV. Prediction Results of LSTM-WQI-BP Model for Various Epochs

Dataset	Epochs	MAE	MSE	RMSE	R2 Score
WQI-BP	20	0.591	0.613	0.7829	0.592
	50	0.585	0.596	0.7720	0.604
	100	0.563	0.583	0.7635	0.627
	150	0.552	0.547	0.7396	0.636
	200	0.543	0.535	0.7314	0.657
	500	0.524	0.513	0.7162	0.684

The performance of the GRU-based WQI prediction model (GRU-WQI-BP model) is assessed by conducting experiments with a range of epochs, varying from 20 to 500. At 20 epochs, the model exhibits specific performance metrics, including a MAE of 0.612, MSE of 0.617, RMSE of 0.7855, and an R2 score of 0.591. As the training advances to 50 epochs, the model demonstrates incremental improvement, leading to lower MAE of 0.605, MSE of 0.59, and RMSE of 0.7681, accompanied by a higher R2 score of 0.618. Subsequently, with 100 epochs, the model's performance continues to enhance, resulting in a reduced MAE of 0.572, while maintaining a similar MSE of 0.571 and exhibiting a lower RMSE of 0.7556 alongside an elevated R2 score of 0.624. The trend of progress is maintained at 150 epochs, where the model showcases further refinement in its predictions, leading to a decreased MAE of 0.564, lower MSE of 0.552, and RMSE of 0.7430, and aR2 score of 0.645. Continuing to 200 epochs, the model consistently improves its predictive accuracy, as reflected by a decreased MAE of 0.547, lower MSE of 0.531, and RMSE of 0.7287, coupled with an increased R2 score of 0.651. Finally, at 500 epochs, the model attains a notable level of predictive excellence, with reduced MAE of 0.536 and MSE of 0.527, further showcasing a lower RMSE of 0.7259, culminating in an impressive R2 score of 0.672. The performance of the GRU-WQI-BP prediction with various epochs is shown in Table XXXV.

Table XXXV. Prediction Results of GRU-WQI-BP Model for Various Epochs

Dataset	Epochs	MAE	MSE	RMSE	R2 Score
WQI-BP	20	0.612	0.617	0.7855	0.591
	50	0.605	0.59	0.7681	0.618
	100	0.572	0.571	0.7556	0.624
	150	0.564	0.552	0.7430	0.645
	200	0.547	0.531	0.7287	0.651
	500	0.536	0.527	0.7259	0.672

In the case of TFT based WQI model, the hyperparameter setting for TFT forecasting involved exploring a prediction time step of 30, setting the encoder layer to 4, using a fixed batch size of 64, varying state sizes from 32 to 256, and setting it to 64, trying out learning rates from 0.0001 to 0.1, varying the number of attention heads from 1 to 8, applying dropout rates from 0 to 0.4. The hyperparameter settings of the TFT-WQI-BP model is tabulated in Table XXXVI.

Table XXXVI: Setting of Special Hyperparameters for TFT Training

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 performance of the TFT-based WQI prediction model (TFT-WQI-BP model) is assessed by conducting experiments with a range of epochs, varying from 20 to 500. At 20 epochs, the model exhibits specific performance metrics, including a MAE of 0.532, MSE of 0.516, RMSE of 0.7183, and an R2 score of 0.623. As the training progresses to 50 epochs, the model showcases incremental refinement, manifesting in reduced MAE of 0.518, lower MSE of 0.498, and a decreased RMSE of 0.7057, accompanied by an elevated R2 score of 0.635. Subsequently, at 100 epochs, the model performance continues to advance, leading to a further reduced MAE of 0.507, with corresponding decreases in both MSE of 0.482 and RMSE of 0.6943, ultimately culminating in an enhanced R2 score of 0.646. As the model is further trained with 150 epochs, it continues to demonstrate improved predictive accuracy, reflected in the reduction of MAE of 0.481, lower MSE of 0.465, and a diminished RMSE of 0.6819, along with an elevated R2 score of 0.662. Continuing the training process to 200 epochs, the model consistently enhances its predictive capability, resulting in a notable decrease in MAE of 0.448, MSE of 0.451, and RMSE of 0.6716, along with a substantial increase in R2 score of 0.687. Finally, with 500 epochs, the model achieves a commendable level of predictive excellence, as evidenced by the decreased MAE of 0.407 and MSE of 0.436, coupled with a further reduction in RMSE 0.6603, culminating in an impressive R2 score of 0.705. The performance of the TFT-WQI-BP prediction with various epochs is shown in Table XXXVII.

Table XXXVII. Prediction Results of TFT-WQI-BP Model for Various Epochs

Dataset	**Epochs**	**MAE**	**MSE**	**RMSE**	**R2 Score**
WQI-BP	20	0.532	0.516	0.7183	0.623
	50	0.518	0.498	0.7057	0.635
	100	0.507	0.482	0.6943	0.646
	150	0.481	0.465	0.6819	0.662
	200	0.448	0.451	0.6716	0.687
	500	0.407	0.436	0.6603	0.705

The prediction results of WQI models for various epochs and dropouts have been observed while implementing deep learning algorithms to discover the best prediction results. It is proved that the models trained with 500 epochs with other hyperparameters such as adam optimizer, momentum as 0.8, dropout as 0.3 and activation function as relu for RNN, LSTM, GRU, and architecture TFT, produced the best results and are shown in Table XXXVIII and depicted in Fig. 7.2.

Table XXXVIII. Overall Performance Results of WQI Models for Bharathapuzha Data

Dataset	Epoch	Models	MAE	MSE	RMSE	R2 Score
WQI-BP	500	RNN-WQI-BP	0.598	0.523	0.7232	0.642
		LSTM-WQI-BP	0.524	0.513	0.7162	0.684
		GRU-WQI-BP	0.536	0.527	0.7259	0.672
		TFT-WQI-BP	**0.407**	**0.436**	**0.6603**	**0.705**

Fig.7.2. Prediction Performance of Deep Learning Algorithms with Bharathapuzha Data

From the above results, it is observed that the TFT-based WQI prediction model shows promising results with a high R2 score value and less error rate. The mean absolute error for TFT based forecasting model is found less as compared to RNN, LSTM and GRU algorithms. The root mean squared error is observed to be less for the TFT-WQI-BP model when compared with RNN-WQI-BP, LSTM-WQI-BP and GRU-WQI-BP prediction model results. The R2 score value defines the accuracy of the model and is observed to be high for the TFT-WQI-BP forecasting model compared with other prediction models.

The prediction results of WQI models trained with the WQI-BP dataset containing a smaller number of instances is compared with the prediction results of WQI models trained with the WQI-SA dataset. It is found that WQI model built with WQI-BP dataset yield less results with respect to metrics. The comparison results of WQI models built with two datasets is shown in Table XXXIX.

Table XXXIX. Comparative Prediction Results of WQI Models Bhavani and Bharathapuzha River Data

Dataset	Models	MAE	MSE	RMSE	R2 Score
WQI-SA	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
	TFT-WQI-SA	**0.122**	**0.167**	**0.4087**	**0.941**
WQI-BP	RNN-WQI-BP	0.598	0.523	0.7232	0.642
	LSTM-WQI-BP	0.524	0.513	0.7162	0.684
	GRU-WQI-BP	0.536	0.527	0.7259	0.672
	TFT-WQI-BP	0.407	0.436	0.6603	0.705

Hence it is required to enhance the performance of WQI models for the Bharathapuzha river and construct a generalized WQI model. For this transfer learning approach is adopted, which are described in the following sections.

7.2. WQI PREDICTION MODELS USING LSTM PRE-TRAINED MODEL

The main focus of this work is to enhance the performance of WQI prediction models for the Bharathapuzha River using transfer learning and deep learning architectures. Transfer learning is shown to be a valuable tool in improving the accuracy of water quality predictions, as it allows models to leverage knowledge from pre-trained models and adapt it to real-world river water quality datasets. The WQI prediction models developed using the Bhavani River WQI-SA dataset, described in chapter 5 is used as a pre-trained model to gain the water quality knowledge of the Bhavani River, which had a larger number of training instances. The study utilizes various deep learning architectures, such as RNN, LSTM and GRU for training and model building. The Temporal Fusion Transformer (TFT) represents a distinct architectural paradigm, which has not been included within the scope of this study and is included in the next section i.e., section 7.3. Through this approach, a generalized framework is suggested to build a robust WQI prediction model for any river. The methodology of building the WQI prediction models using transfer learning the WQI-BP dataset is described below.

Methodology

Transfer learning approach offer significant benefits by leveraging knowledge gained from pre-trained models on large datasets, allowing them to perform well even with limited training data for specific tasks. This approach enables faster convergence, reduces computational costs, and enhances prediction accuracy, making it particularly advantageous in various real-world applications. Here in this work as there is no existing pre-trained WQI model for time series data, the developed model approach is applied.

The proposed transfer learning based WQI prediction model is constructed using three essential components such as (i) training dataset (ii) pre-trained WQI model (iii) method for transfer learning. The target variable for regression modelling is the water quality index. The transfer learning-based WQI prediction models are built using RNN, LSTM, and GRU algorithms, using the WQI-BP dataset and pre-trained model developed in Chapter 5 as shown in the framework Fig. 7.3.

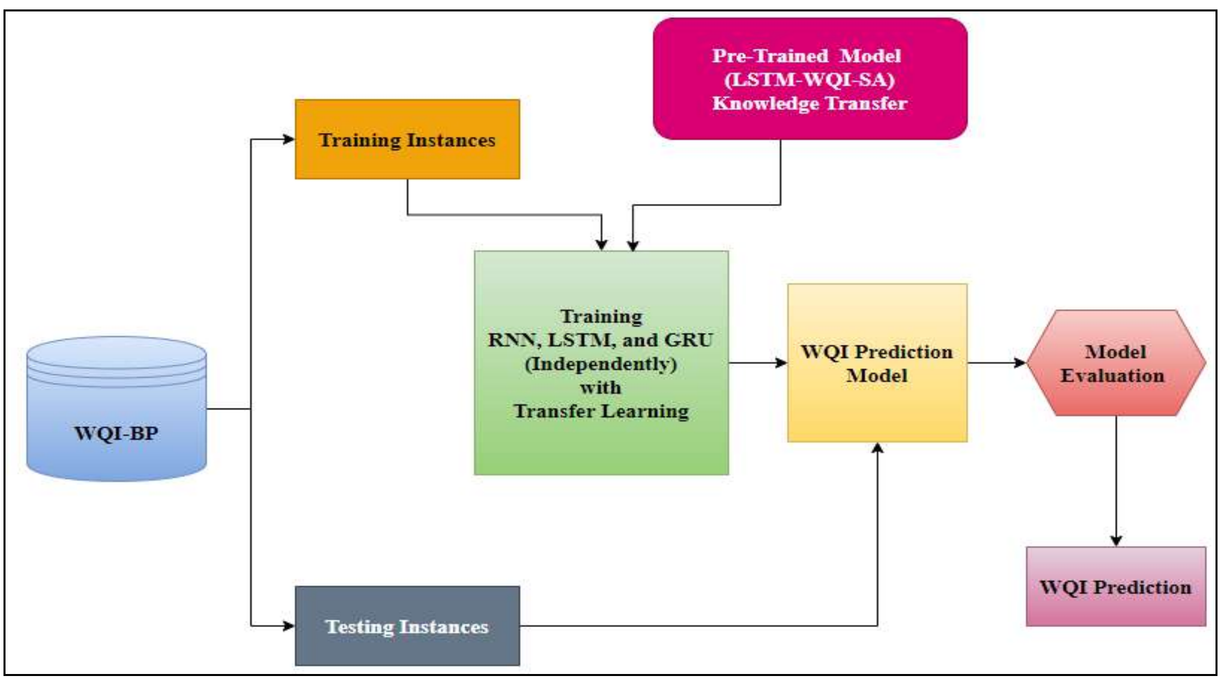

Fig. 7.3. The Methodology of Homogenous Transfer Learning based WQI Prediction Model

Training Dataset

The river water quality samples consisting of 26 physicochemical parameters are collected from three monitoring stations of the Bharathapuzha River and 10 corresponding seasonal characteristics recorded at visual crossing sites are derived. Also, longitude, latitude, station ID, and date have been included. The water quality index for each sample is determined based on the Indian Standard for Drinking Water Specification and augmented as a target variable. The time series data with 2190 samples is developed.

Exploratory data analysis is performed on the raw data to better understand the characteristics of the data and the importance of each parameter in determining the water quality index. The techniques, such as heatmap analysis, boxplot analysis, pair plot analysis, and histogram analysis, are used for analysis. EDA suggested the need for data pre-processing, including min-max normalization and data cleaning. The select K best feature selection method is used, resulting in the creation of the WQI-BP dataset, which consists of 2190 instances and 39 attributes, as outlined in Table XII of Chapter 3.

Pre-trained Model

The WQI models developed using Bhavani River data and trained with RNN, LSTM and GRU as described in Chapter 5 are considered here as pre-built models. The optimal performance of WQI models has been achieved when trained with 500 epochs and a dropout rate of 0.3. Notably, LSTM-WQI-SA models demonstrated superior R2 scores and lower error rates compared to RNN-WQI-SA and GRU-WQI-SA models as per the results given in Table XXVI, which is reproduced in Table XL.

Table XL. Prediction Results of Pre-trained Model with WQI-SA Dataset

Dataset	Dropout	Epoch	Models	MAE	MSE	RMSE	R2 Score
WQI-SA	0.3	500	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

The LSTM-WQI-SA model developed for predicting water quality index yielded better results and this model is utilized as the pre-trained model for transfer learning. The weights of the parameters can be transferred while training Bharathapuzha River water data using transfer learning to achieve a better prediction model. This approach enables the adaptation of the pre-trained model to different geographical regions, which do not have sufficient data for training a model from scratch.

Model Building Using Homogenous Transfer Learning

Transfer learning is a powerful technique in machine learning that allows a model to leverage knowledge gained from one domain and apply it to another related domain. Transfer learning can greatly reduce the amount of labelled training data required, as the pre-trained model already possesses a significant amount of knowledge from the source domain. Here the set of parameters considered for Bharathapuzha river data is the same as the set of attributes considered for building pre-trained model. Hence the transfer learning approach is homogenous transfer learning.

Here again, the deep learning architectures RNNs, LSTMs, and GRUs each offering unique advantages in handling sequential data, and are trained independently using the WQI-BP dataset.

Training in transfer learning with deep learning architectures involves a multi-step process that capitalizes on pre-existing knowledge and fine-tuning to enhance the performance of a model on a specific task, such as the WQI prediction using the WQI-BP dataset. Firstly, a pre-trained model, often trained on a larger and related dataset, is selected. In this case, LSTM based pretrained WQI model built with Bhavani River data is used as pre-trained model. This pre-trained model serves as a feature extractor, capturing general patterns applicable to time series data. Next, the chosen pre-trained model is adapted to the WQI-BP dataset through a process of fine-tuning. This involves modifying the architecture by adding additional layers, adjusting weights, and configuring hyperparameters to tailor the model for the specific task of predicting water quality. The model architecture retains its learned features while being further trained on the WQI-BP dataset, allowing it to capture domain-specific patterns and nuances.

The 80% of instances of the WQI-BP dataset prepared are given as input along with knowledge taken from the pre-trained model (LSTM-WQI-SA model) to RNN and its variants LSTM and GRU for training the networks.

The best hyperparameters are chosen during model training to make the model more effective mapping the input features as independent variables to the target variable as the dependent variable. In designing and training RNNs, LSTMs, and GRUs, several crucial hyperparameters are employed to optimize their performance and generalization. The learning rate is a fundamental hyperparameter shared across all these models, determining the step size taken during gradient descent to update model weights. Setting an appropriate learning rate is essential to avoid slow convergence or overshooting. Additionally, the number of hidden units controls the dimensionality of the hidden state and memory cells in LSTM and GRU. The regularization techniques, such as dropout are applied to prevent overfitting. Finally, the WQI prediction models are built by learning water quality patterns from the input instances of the WQI-BP dataset and the knowledge learned from the pre-trained model using RNN, LSTM, and GRU with proper hyperparameters settings and the new models are referred as RNN-WQI-BP-TL, LSTM-WQI-BP-TL and GRU-WQI-BP-TL.

Evaluation metrics such as MAE, MSE, RMSE, and R2 score are used to measure the performance of predictive models using 20% of the instances in the WQI-BP dataset as test set.

Experiments and Results

The experiments have been carried out by training the WQI-BP dataset and the pre-trained LSTM-WQI-SA model using deep learning algorithms such as RNN, LSTM and GRU and implemented independently using Python libraries under TensorFlow, Keras and scikit learn. The training dataset contains 1752 tagged instances of the WQI-BP dataset. Evaluation of the prediction models is carried out to check the efficiency of the model using the metrics such as MAE, MSE, RMSE and R2 score values with the test data set containing 438 instances.

The deep learning architectures RNN, LSTM, and GRU, are set with carefully defined hyperparameters. The dense layers are configured with unit values ranging from 5 to 10, and the optimizer utilized is the adam optimizer. The training process involves various epoch sizes, specifically 20, 50, 100, 150, 200, and 500. For activation functions, ReLU is chosen, while the momentum parameter is set within the range of 0.5 to 0.9. A dropout rate of 0.2 is applied, and the learning rate is set to 0.1, with the batch size varying between 32 and 64. A summary of the hyperparameter settings for training the deep neural networks is presented in Table XLI. The experimental results demonstrate that employing a momentum value of 0.8, an epoch size of 500, a dropout rate of 0.3, and using the ReLU activation function yields superior outcomes.

Table XLI. Hyperparameters Setting for Training RNN, LSTM, GRU

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

The RNN and transfer learning based WQI prediction model (RNN-WQI-BP-TL model) trained using LSTM-WQI-SA pretrained model and WQI-BP dataset have been experimented with various epochs such as from 20 to 500 where various metrics are measured at different epochs. Starting at 20 epochs, the MAE is calculated at 0.526, illustrating the average absolute difference between predicted and actual values. Concurrently, the MSE is measured as 0.517, indicating the mean of squared differences. The RMSE value is observed at 0.7190, reflecting a square root of the MSE. The R2 score, denoting the coefficient of determination, is reported at 0.798. Advancing to 50 epochs, a decrease in MAE is evident, measuring 0.473, indicating a smaller average absolute difference. The MSE also decreases to 0.475, reflecting reduced squared differences. Correspondingly, the RMSE value diminishes to 0.6892, while the R2 score ascends to 0.816, showcasing an improved predictive capability. At 100 epochs, a further decrease in MAE is observed, reaching 0.423, reflecting enhanced accuracy. This trend extends to the MSE, which decreases to 0.427, and the RMSE, which reduces to 0.6535. The R2 score continues to climb to 0.82, indicating heightened predictive precision.

Moving to 150 epochs, both MAE and MSE maintain relatively steady values at 0.4 and 0.401, respectively. The RMSE value remains consistent at 0.6332. The R2 score progresses to

0.83, implying an improved level of goodness of fit between predictions and actual data. As the epoch count advances to 200, both MAE and MSE retain similar values at 0.398 and 0.391, respectively. The RMSE value also sustains its level at 0.6253. The R2 score maintains a stable value of 0.836, indicating a consistent predictive capability. Upon reaching 500 epochs, the dataset demonstrates the lowest MAE of 0.371, reflecting highly accurate predictions. The MSE, RMSE, and R2 Score are reported at 0.387, 0.6221, and 0.84 respectively. The performance results of the RNN-based WQI prediction model across different epochs are tabulated in Table XLII.

Table XLII. Results of RNN and TL based WQI Models for Different Epochs

Dataset	Epochs	MAE	MSE	RMSE	R2 Score
WQI-BP	500	0.371	0.387	0.6221	0.84
	200	0.398	0.391	0.6253	0.836
	150	0.4	0.401	0.6332	0.83
	100	0.423	0.427	0.6535	0.82
	50	0.473	0.475	0.6892	0.816
	20	0.526	0.517	0.7190	0.798

The LSTM and transfer learning based WQI prediction model (LSTM-WQI-BP-TL model) trained using LSTM-WQI-SA pretrained model and WQI-BP dataset have been experimented with various epochs such as from 20 to 500 where various metrics are measured at different epochs. Starting with 20 epochs, the MAE is noted at 0.322, indicating a modest average absolute difference between predicted and actual values. The MSE is observed at 0.354, signifying the mean of squared differences. The RMSE value is reported as 0.5950, reflecting the square root of the MSE. The R2 score, denoting the coefficient of determination, stands at 0.8, suggesting a high proportion of variance in the dependent variable that can be explained by independent variables. Advancing to 50 epochs, the MAE experiences a slight decrease to 0.316, indicative of improved predictive precision. This trend is echoed in the MSE, which diminishes to 0.325, and the RMSE, which reduces to 0.5701. The R2 score continues to climb to 0.825, denoting an enhanced goodness of fit between predictions and actual data. As the epoch count reaches 100, the MAE further reduces to 0.298, indicating refined predictive accuracy. The MSE also decreases to 0.312, reflecting reduced squared differences. The RMSE value diminishes to 0.5586, while the R2 Score decreases slightly to 0.8321, denoting a stable predictive capability.

Progressing to 150 epochs, both MAE and MSE maintain steady values at 0.292 and 0.263, respectively. The RMSE remains consistent at 0.5128. The R2 score increases to 0.841, highlighting an improved level of goodness of fit between predictions and actual data. Upon reaching 200 epochs, the MAE experiences a substantial drop to 0.285, reflecting heightened predictive accuracy. The MSE and RMSE values further decrease to 0.182 and 0.4266, respectively. The R2 score climbs to 0.862, signalling an enhanced predictive capability. At 500 epochs, the dataset demonstrates the lowest MAE of 0.229, signifying a high degree of predictive accuracy. The MSE, RMSE, and R2 score are reported at 0.163, 0.4037, and 0.87 respectively. The performance results of the LSTM-based WQI prediction model across different epochs are tabulated in Table XLIII.

Table XLIII. Results of LSTM and TL based WQI Models for Different Epochs

Dataset	Epochs	MAE	MSE	RMSE	R2 Score
WQI-BP	500	0.229	0.163	0.4037	0.87
	200	0.285	0.182	0.4266	0.862
	150	0.292	0.263	0.5128	0.841
	100	0.298	0.312	0.5586	0.8321
	50	0.316	0.325	0.5701	0.825
	20	0.322	0.354	0.5950	0.8

The results of the GRU and transfer learning-based WQI prediction model (GRU-WQI-BP-TL model) trained using the LSTM-WQI-SA pretrained model and WQI-BP dataset have been experimented with various epochs such as from 20 to 500 where various metrics are measured at different epochs. Commencing with 20 epochs, the MAE stands at 0.372, indicating a moderate average absolute difference between predicted and actual values. The MSE is recorded as 0.387, representing the mean of squared differences. The RMSE is reported at 0.6221, signifying the square root of the MSE. The R2 Score, representing the coefficient of determination, is noted as 0.813, denoting the proportion of variance in the dependent variable that can be explained by the independent variables. Advancing to 50 epochs, a reduction in MAE is observed, reaching 0.337, indicating an improvement in predictive accuracy. The MSE value also decreases to 0.341, reflecting a reduction in squared differences. Correspondingly, the RMSE value diminishes to 0.5840, while the R2 score ascends to 0.821, highlighting enhanced goodness of fit between predictions and actual data. At 100 epochs, the MAE experiences a further reduction to 0.324, signifying an improvement in predictive precision. The MSE and RMSE values both decrease to 0.326 and 0.5710 respectively, while the R2 score continues to climb to 0.837, indicating an advanced level of predictive capability.

Moving to 150 epochs, the MAE and MSE maintain consistent values at 0.316 and 0.318 respectively, reflecting steady predictive accuracy. The RMSE value remains relatively stable at 0.5639. The R2 score further progresses to 0.843, denoting an enhanced level of goodness of fit between predictions and actual data. Upon reaching 200 epochs, the dataset reveals a decrease in MAE to 0.298, reflecting heightened predictive accuracy. Similarly, the MSE and RMSE values decrease to 0.296 and 0.5441 respectively. The R2 score advances to 0.851, indicating an improved predictive capability. Finally, at 500 epochs, the dataset demonstrates the lowest MAE of 0.282, signifying a high degree of predictive accuracy. The MSE, RMSE, and R2 score are reported at 0.273, 0.5225, and 0.862 respectively. The performance results of the GRU-based WQI prediction model across different epochs are tabulated in Table XLIV.

Table XLIV. Results of GRU and TL based WQI Models for Different Epochs

Dataset	Epochs	MAE	MSE	RMSE	R2 Score
WQI-BP	500	0.282	0.273	0.5225	0.862
	200	0.298	0.296	0.5441	0.851
	150	0.316	0.318	0.5639	0.843
	100	0.324	0.326	0.5710	0.837
	50	0.337	0.341	0.5840	0.821
	20	0.372	0.387	0.6221	0.813

The performance results of the RNN-WQI-BP-TL model for different dropout rates are tabulated in Table XLV. The provided dataset offers a comprehensive comparison of various models and dropout rates with respect to their predictive performance on the WQI-BP dataset. For the RNN-WQI-BP-TL model, employing a dropout rate of 0.3 yields MAE of 0.371, MSE of 0.387, RMSE of 0.6221, and an R2 Score of 0.84. Similarly, at a dropout rate of 0.2, the model's performance is slightly reduced, resulting in a MAE of 0.4, MSE of 0.401, RMSE of 0.6332, and R2 Score of 0.83. Switching to the LSTM-WQI-BP-TL model, a dropout rate of 0.3 leads to an improved performance, with an MAE of 0.229, MSE of 0.163, RMSE of 0.4037, and R2 Score of 0.87.

At a dropout rate of 0.2, the model maintains favourable predictive accuracy, resulting in a MAE of 0.292, MSE of 0.263, RMSE of 0.5128, and R2 Score of 0.841. For the GRU-WQI-BP-TL model, a dropout rate of 0.3 showcases competitive performance, with an MAE of 0.282, MSE of 0.273, RMSE of 0.5225, and R2 Score of 0.862. Similarly, at a dropout rate of 0.2, the model's performance remains favourable, resulting in an MAE of 0.316, MSE of 0.318, RMSE of 0.5639, and R2 Score of 0.843. These findings highlight the impact of dropout rates on the performance of the RNN-WQI-BP-TL model, as reflected in the variations of MAE, MSE, RMSE, and R2 score.

Table XLV. Results of Different TL based WQI Prediction Models for Different Dropout Rates

Dataset	Models	Dropouts	MAE	MSE	RMSE	R2 Score
WQI-BP	RNN-WQI-BP-TL	0.3	0.371	0.387	0.6221	0.84
		0.2	0.4	0.401	0.6332	0.83
	LSTM-WQI-BP-TL	0.3	0.229	0.163	0.4037	0.87
		0.2	0.292	0.263	0.5128	0.841
	GRU-WQI-BP-TL	0.3	0.282	0.273	0.5225	0.862
		0.2	0.316	0.318	0.5639	0.843

The prediction results of WQI prediction models built using the knowledge gained from LSTM-WQI-SA model and the Bharathapuzha River data (WQI-BP) for various epochs and dropouts have been observed while implementing deep learning algorithms to discover the best prediction results. It is proved that the models trained with 500 epochs and dropout rate 0.3 with other hyperparameters such as adam optimizer, momentum as 0.8 and activation function as relu for RNN, LSTM and GRU produced the best results and are shown in Table XLVI and depicted in Fig. 7.4.

Table XLVI. Final Prediction Results of Various TL based WQI Models with LSTM Pretrained Model

Dataset	Dropout	Epoch	Models	MAE	MSE	RMSE	R2 Score
WQI-BP	0.3	500	RNN-WQI-BP-TL	0.371	0.387	0.6221	0.84
			LSTM-WQI-BP-TL	**0.229**	**0.163**	**0.4037**	**0.87**
			GRU-WQI-BP-TL	0.282	0.273	0.5225	0.862

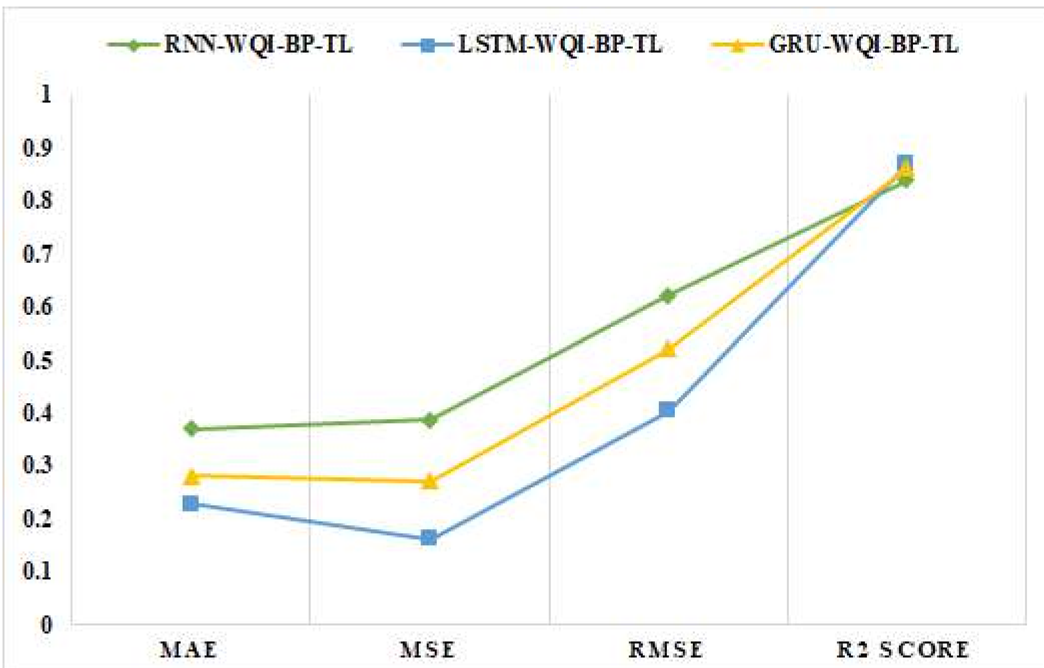

Fig.7.4. Final Prediction Results of Various TL based WQI Models with LSTM Pretrained Model

From the above results, it is observed that the LSTM-WQI-BP-TL, WQI prediction model shows promising results with a high R2 score value and less error rate. The mean absolute error for LSTM based forecasting model is found less as compared to RNN and GRU algorithms. The root mean squared error is observed to be less for the LSTM-WQI-BP-TL model when compared with RNN-WQI-BP-TL and GRU-WQI-BP-TL prediction model results. The R2 score value defines the accuracy of the model and is observed to be high for the LSTM-WQI-BP-TL forecasting model compared with other prediction models.

Comparative Analysis of TL-based WQI Prediction Models Verus Base WQI Models

The performance results of the TL based WQI prediction models built using the pre-trained model (LSTM-WQI-SA) and WQI-BP dataset is compared with the prediction results of WQI models built with limited instances of Bharathapuzha river data

While training under a dropout rate of 0.3 and after 500 epochs, the RNN-WQI-BP-TL model showcases a MAE of 0.371, a MSE of 0.387, a RMSE of 0.6221, and an R2 score of 0.84. In the same configuration, the LSTM-WQI-BP-TL model demonstrates superior predictive performance with an MAE of 0.229, MSE of 0.163, RMSE of 0.4037, and an impressive R2 score of 0.87. Similarly, the GRU-WQI-BP-TL model, works with the same dropout rate and epoch count, exhibits notable accuracy, with an MAE of 0.282, MSE of 0.273, RMSE of 0.5225, and an R2 score of 0.862.

The comparison also extends to other models built without using transfer learning approaches. For instance, RNN-WQI-BP records a MAE of 0.598, MSE of 0.523, RMSE of 0.7232, and an R2 score of 0.642. In contrast, LSTM-WQI-BP yields improved results with a MAE of 0.524, MSE of 0.513, RMSE of 0.7162, and an R2 score of 0.684, while GRU-WQI-BP strikes a balance with a MAE of 0.536, MSE of 0.527, RMSE of 0.7259, and an R2 score of 0.672. Finally, TFT-WQI-BP exhibits promising predictive accuracy, featuring a MAE of 0.407, MSE of 0.436, RMSE of 0.6603, and an R2 score of 0.705.

The comparative results of prediction models built with and without transfer learning are tabulated in Table XLVII and illustrated in Fig.7.5. From the prediction results it is proved that the prediction models built with transfer learning give better results. The LSTM-WQI-BP-TL model demonstrated superior performance results in predicting the water quality index, outperforming the other models in terms of MAE, MSE, RMSE, and R2 score. The inclusion of transfer learning appeared to contribute to the enhanced predictive capabilities of the LSTM model. These results suggest that LSTM-WQI-BP-TL shows a promising choice for water quality index prediction.

Table XLVII. Comparative Analysis of TL Based WQI Prediction Models Vs Base Models

Dataset	Dropout	Epoch	Models	MAE	MSE	RMSE	R2 Score
WQI-BP	0.3	500	RNN-WQI-BP-TL	0.371	0.387	0.6221	0.84
			LSTM-WQI-BP-TL	**0.229**	**0.163**	**0.4037**	**0.87**
			GRU-WQI-BP-TL	0.282	0.273	0.5225	0.862
			RNN-WQI-BP	0.598	0.523	0.7232	0.642
			LSTM-WQI-BP	0.524	0.513	0.7162	0.684
			GRU-WQI-BP	0.536	0.527	0.7259	0.672
			TFT-WQI-BP	0.407	0.436	0.6603	0.705

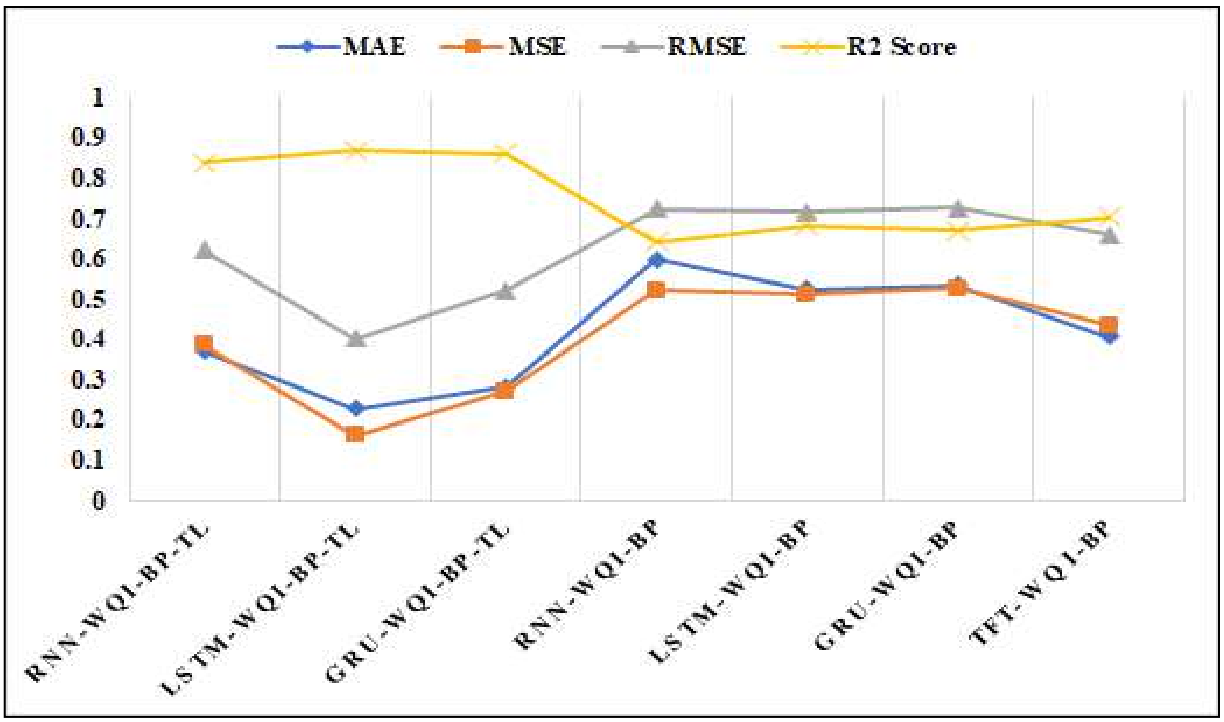

Fig.7.5. Comparative Analysis of TL-Based WQI Prediction Models Vs Base Models

Findings

Transfer learning enhances the performance of deep neural network models by significantly improving the accuracy and predictive power in predicting river water quality index. The transfer learning in formulating a WQI prediction model tailored to the Bharathapuzha River has yielded promising outcomes. Through the strategic utilization of a pre-trained model generated from the extensive WQI-SA dataset, which encompassed a substantial array of physiochemical and seasonal parameters derived from the Bhavani River over a five-year span, an adept prediction model was successfully devised, even in the face of constrained time series data inherent to the Bharathapuzha River. By comparing models trained from scratch and models fine-tuned with transfer learning, the study demonstrates the effectiveness of transfer learning in contributing to the field of water quality analysis and improving the performance of deep neural network models for WQI prediction. This revelation underscores the potency of transfer learning as a valuable asset in constructing precise and streamlined prediction models for the evaluation of water quality, particularly in regions marked by limited data availability. Transfer learning helps in improving the generalization and robustness of DNN models in river WQI prediction.

7.3 WQI PREDICTION MODELS USING TFT PRE-TRAINED MODEL

The main aim of this work is to enhance the efficacy of the water quality prediction model applied to the Bharathapuzha River. Here the enhancement is achieved through the implementation transfer learning technique with TFT based pretrained model. This work also employs a spectrum of deep learning architectures including RNN, LSTM, GRU and the TFT architecture. The reason for using again RNN, LSTM, GRU simultaneously when the TFT architecture is employed for transfer learning-based training is to study and confirm which deep neural network performs better when the transfer learning approach is adopted. The WQI prediction model developed using the Bhavani River data i.e., WQI-SA dataset trained with TFT, described in Chapter 6 is used as a pre-trained model to gain the water quality knowledge of the Bhavani River, which has a larger number of training instances. Through this approach, a generalized framework is suggested to build a robust WQI prediction model for any river.

Methodology

TFT represent a significant advancement in the field of sequential modelling and forecasting. This specialized neural network architecture builds upon the self-attention mechanism of transformer model by introducing temporal fusion, which allows them to effectively capture both temporal dependencies and dynamic patterns in time-series data. Transfer learning approach offer significant benefits by leveraging knowledge gained from pre-trained models on large datasets, allowing them to perform well even with limited training data for specific tasks. Here in this work as there is no existing pre-trained model for time series water quality prediction, the developed model approach is applied.

The proposed transfer learning-based WQI prediction model is constructed using three essential components such as (i) training dataset (ii) pre-trained WQI model (iii) networks for transfer learning. The transfer learning-based WQI prediction models are built by training the networks RNN, LSTM, GRU and also training TFT independently, as shown in the framework Fig. 7.6 using the WQI-BP dataset and the TFT based pre-trained model developed in Chapter 6.

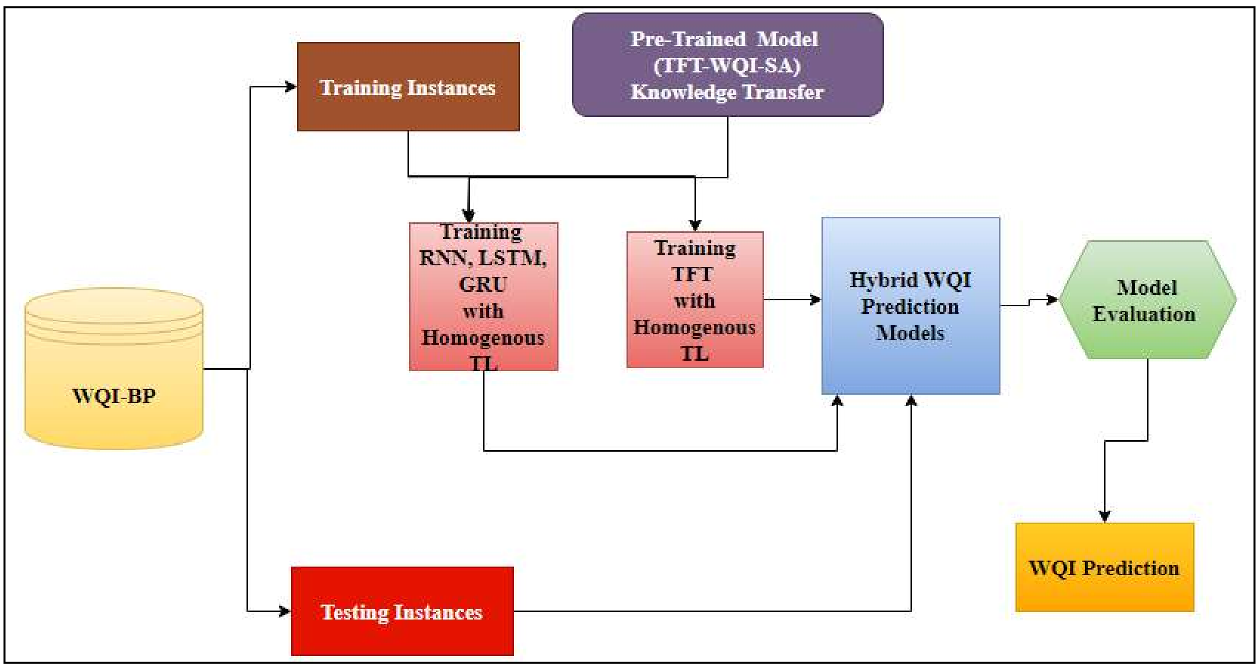

Fig. 7.6. The Methodology of Transfer Learning based TFT Pre-trained Model

Training Dataset

Twenty-six physicochemical parameters derived from three monitoring stations along the Bharathapuzha River are coupled with 10 seasonal characteristics documented at visual crossing sites, and with essential geographical attributes such as longitude, latitude, station ID, and date for preparing the time series river water samples. The WQI is calculated and augmented with corresponding instances. The time series data with 41 attributes and 2190 instances forms a robust foundation for the development of WQI prediction models. Through an exploratory data analysis undertaken on the raw data, a comprehensive comprehension of data characteristics and parameter significance in determining the water quality index is attained. Employing techniques such as heatmap analysis, boxplot analysis, pair plot analysis, and histogram analysis, the EDA underscores the necessity for data pre-processing, encompassing min-max normalization and data cleansing. The application of the select K best feature selection method leads to the creation of the WQI-BP dataset, comprising 2190 instances and 38 attributes, which was presented in detail in Chapter 3 and listed in Table XII.

Pre-trained Model

The WQI models developed using Bhavani River data i.e., the WQI-SA dataset and trained with the TFT a sophisticated architecture for time series, as described in Chapter 6 is considered here as a pre-trained model. The optimal performance of TFT based WQI model has been achieved when trained with 500 epochs and a dropout rate of 0.3. The TFT-WQI-SA model demonstrated superior R2 scores and lower error rates as per the results given in Table. XXXII, which is reproduced in Table XLVIII.

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

Model	**Dropout**	**Epoch**	**MAE**	**MSE**	**RMSE**	**R2 Score**
RNN -WQI-SA	**0.3**	**500**	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
TFT-WQI-SA			**0.122**	**0.167**	**0.4087**	**0.941**

Hence the TFT-WQI-SA model, developed for the prediction of water quality index, is used as the pre-trained model in this work to facilitate transfer learning. The weights of the parameters can be transferred while training Bharathapuzha River water data using transfer learning to achieve a better prediction model.

Model Building using Homogenous Transfer Learning

Transfer learning accelerates the training process significantly since the model commences with higher accuracy and generalization capabilities. By leveraging pre-trained models with substantial knowledge from a source domain, transfer learning substantially reduces the need for labelled training data.

Here, to build transfer learning based WQI prediction models, the networks RNN, LSTM, GRU and TFT are trained independently. The process of transfer learning in the context of the WQI prediction model involves two distinct steps, designed to leverage knowledge from one river data to enhance predictions in another river data. Initially, the pre-trained model, TFT-WQI-SA, is employed as the starting point. To retain the knowledge that the model's layers have acquired, all layers are frozen except the output layer. By maintaining these layers unaltered, the model preserves the knowledge and insights gathered from the Bhavani River data, where it was initially trained. This preservation of knowledge proves advantageous for predicting the WQI for the Bharathapuzha River, as the underlying factors affecting WQI might share similarities between the two regions.

Subsequently, the process advances to the retraining phase, wherein the frozen model is fine-tuned using the WQI-BP dataset. This step entails adjusting the model's parameters based on the specific attributes of the Bharathapuzha River. By doing so, the model becomes specialized in predicting WQI, capitalizing on the pre-existing understanding transferred from the TFT-WQI-SA model. As a result, the model's predictions for WQI values with the WQI-BP data set are expected to exhibit improved accuracy and efficiency. Essentially, this approach capitalizes on the underlying similarities between the two rivers such as Bhavani and Bharathapuzha rivers, allowing the model to generalize its knowledge and predictions across different rivers.

Strategic adjustments to hyperparameters play a pivotal role in refining the transfer learning process. The performance of the model is significantly improved by carefully tuning hyperparameters such as the learning rate, batch size, number of epochs, activation functions, and regularization strength, The learning rate alteration governs the pace at which the model converges, ensuring an efficient and effective learning trajectory. Batch size manipulation affects weight updates, with smaller batches potentially aiding the model escape from local optima and enhancing its ability to generalize. The number of epochs adjustment regulates the depth of training, striking a balance between underfitting and overfitting. The choice of activation functions influences the capacity of the model to capture intricate patterns and gradients. Meanwhile, tuning the regularization strength, exemplified by L1 or L2 regularization, helps counteract overfitting by controlling the magnitude of model weights.

In case of training the TFT network for homogenous transfer learning, the special hyperparameters such as time step, the encoder layer, batch sizes, state sizes, learning rates, number of attention heads, dropout rates, and Loss Function values are configured appropriately while training the TFT. This meticulous parameter tuning within the TFT training process is paramount to achieving effective and adaptable knowledge transfer across different domains.

Four different architectures, namely RNN, LSTM, GRU and the TFT, are trained and four independent TL based WQI prediction models are built. The models are named as RNN-WQI-BP-TFL, LSTM-WQI-BP-TFL, GRU-WQI-BP-TFL and TFT-WQI-BP-TFL. Evaluation metrics such as MAE, MSE, RMSE, and R2 scores are used to measure the performance of predictive models using 20% of the instances in the WQI-BP dataset.

Experiments and Results

The experiments have been carried out by training with the WQI-BP dataset and the pre-trained TFT-WQI-SA model and implemented using Python libraries under Tensor. The training dataset contains 1752 tagged instances of the WQI-BP dataset. RNN, LSTM, GRU and the TFT architecture are implemented independently. Evaluation of the prediction models is carried out to check the efficiency of the model using the metrics such as MAE, MSE, RMSE and R2 score with the test data set containing 438 instances. The hyper parameter setting for RNN, LSTM and GRU as mentioned in the previous section with Table XLI.

The RNN and transfer learning based WQI prediction model (RNN-WQI-BP-TFL model) trained using TFT-WQI-SA model and WQI-BP dataset have been experimented with various epochs such as from 20 to 500 where various metrics are measured at different epochs. At 20 epochs, the model achieves an MAE of 0.4, MSE of 0.401, RMSE of 0.6332, and an R2 Score of 0.83. With a higher number of epochs, such as 50, the model demonstrates improvements in its predictive accuracy, yielding an MAE of 0.398, MSE of 0.391, RMSE of 0.6253, and an R2 Score of 0.836. Continuing this trend, the model further refines its predictions at 100 epochs, with an MAE of 0.371, MSE of 0.387, RMSE of 0.6221, and an R2 Score of 0.84. As the model's training progresses, it continues to enhance its performance. At 150 epochs, the MAE decreases to 0.362, MSE to 0.346, RMSE to 0.5882, and the R2 Score improves to 0.846. This trend of improvement is maintained at 200 epochs, with the model achieving an MAE of 0.284, MSE of 0.324, RMSE of 0.5692, and an R2 score of 0.854. Finally, at 500 epochs, the model reaches its highest level of predictive accuracy, yielding an MAE of 0.235, MSE of 0.246, RMSE of 0.4960, and an impressive R2 score of 0.86.

The results indicate that the RNN-WQI-BP-TFL prediction model performs well in predicting the target variable. The performance of RNN-WQI-BP-TFL prediction model is depicted in Table XLIX.

Table XLIX. Performance of RNN-WQI-BP-TFL Model for Different Epochs

Pre-trained Model	Dataset	Epochs	MAE	MSE	RMSE	R2 Score
TFT-WQI-SA	WQI-BP	500	0.235	0.246	0.4960	0.86
		200	0.284	0.324	0.5692	0.854
		150	0.362	0.346	0.5882	0.846
		100	0.371	0.387	0.6221	0.84
		50	0.398	0.391	0.6253	0.836
		20	0.4	0.401	0.6332	0.83

The results of the LSTM and transfer learning based WQI prediction model (LSTM-WQI-BP-TFL model) trained using the TFT-WQI-SA model and WQI-BP dataset have been experimented with various epochs such as from 20 to 500 where various metrics are measured at different epochs. Beginning with 20 epochs, the model exhibits an MAE of 0.298, MSE of 0.291, RMSE of 0.5394, and an R2 Score of 0.841. As the number of epochs increases to 50, the model's predictive accuracy improves, resulting in an MAE of 0.292, MSE of 0.263, RMSE of 0.5128, and an R2 Score of 0.852. Continuing the trend, at 100 epochs, the model's performance further advances, yielding an MAE of 0.285, MSE of 0.182, RMSE of 0.4266, and an R2 score of 0.874. The predictive accuracy continues to improve at 150 epochs, with an MAE of 0.229, MSE of 0.163, RMSE of 0.4037, and an R2 Score of 0.88. Further refinement is observed at 200 epochs, where the model demonstrates enhanced predictive capabilities with an MAE of 0.217, MSE of 0.159, RMSE of 0.3987, and an R2 score of 0.892. Finally, at 500 epochs, the model reaches its peak predictive performance, achieving an MAE of 0.195, MSE of 0.137, RMSE of 0.3701, and an impressive R2 score of 0.902. The performance of LSTM-WQI-BP-TFL prediction model is depicted in Table L.

Table L. Performance of LSTM-WQI-BP-TFL Prediction Model for Different Epochs

Dataset	Epochs	MAE	MSE	RMSE	R2 Score
WQI-BP	500	0.195	0.137	0.3701	0.902
	200	0.217	0.159	0.3987	0.892
	150	0.229	0.163	0.4037	0.88
	100	0.285	0.182	0.4266	0.874
	50	0.292	0.263	0.5128	0.852
	20	0.298	0.291	0.5394	0.841

The results of the GRU and transfer learning-based WQI prediction model (GRU-WQI-BP-TFL model) trained using the TFT-WQI-SA model and WQI-BP dataset have been experimented with various epochs such as from 20 to 500 where various metrics are measured at different epochs. At 20 epochs, the model achieves a MAE of 0.298, a MSE of 0.291, a RMSE of 0.5394, and an R2 Score of 0.841. As the number of training epochs increases to 50, the models predictive accuracy improves, resulting in an MAE of 0.292, MSE of 0.263, RMSE of 0.5128, and an R2 score of 0.852. Continuing along this trajectory, at 100 epochs, the model's performance demonstrates further advancement, yielding an MAE of 0.285, MSE of 0.182, RMSE of 0.4266, and an R2 score of 0.87. As the number of epochs rises to 150, the model predictive capabilities continue to evolve, achieving an MAE of 0.226, MSE of 0.163, RMSE of 0.4037, and an R2 score of 0.875. With a focus on 200 epochs, the model continues to refine its predictive accuracy, delivering an MAE of 0.219, MSE of 0.159, RMSE of 0.3987, and an R2 score of 0.881. Finally, at 500 epochs, the model reaches its zenith in predictive performance, boasting an MAE of 0.215, MSE of 0.137, RMSE of 0.3701, and a commendable R2 score of 0.892. The Performance of GRU-WQI-BP-TFL prediction model is depicted in Table LI.

Table LI. Performance of GRU-WQI-BP-TFL Prediction Model for Different Epochs

Dataset	Epochs	MAE	MSE	RMSE	R2 Score
WQI-BP	500	0.215	0.137	0.3701	0.892
	200	0.219	0.159	0.3987	0.881
	150	0.226	0.163	0.4037	0.875
	100	0.285	0.182	0.4266	0.87
	50	0.292	0.263	0.5128	0.852
	20	0.298	0.291	0.5394	0.841

In case of TFT network implementation the model employs the Adam, optimization. The optimal model selection during training is facilitated through 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. The final parameters for the enhanced TFT of the two units are presented in Table LI.

Table LII. Special Hyperparameters for TFT Training

Time steps	Encoders layers	Batch sizes	State size	Learnin g rates	Attentio n heads	Dropout rate	Loss Function a	Loss Function b	Loss Function g
30	4	64	64	0.01	4	0.2, 0.3	0.80	0.01	0.10

The results of the TFT and TL-based WQI prediction model (TFT-WQI-BP-TFL model) trained using the TFT-WQI-SA model and WQI-BP dataset have been experimented with various epochs such as from 20 to 500 where various metrics are measured at different epochs. At 20 epochs, the model achieves MAE of 0.279, MSE of 0.293, and RMSE of 0.5413. The R2 Score, indicating the goodness of fit, stands at 0.88. Moving to 50 epochs, resulting in an MAE of 0.257, MSE of 0.281, RMSE of 0.5301, and an R2 Score of 0.89.

As training progresses to 100 epochs, the performance of the model demonstrates further enhancement, yielding an MAE of 0.215, MSE of 0.269, RMSE of 0.5187, and an R2 Score of 0.913. Continuing along this trajectory, at 150 epochs, the model predictive capabilities continue to evolve, achieving an MAE of 0.142, MSE of 0.247, RMSE of 0.4970, and an R2 Score of 0.932. With a focus on 200 epochs, the model refines its predictive accuracy even further, delivering an MAE of 0.138, MSE of 0.183, RMSE of 0.4278, and an impressive R2 Score of 0.957. Finally, at 500 epochs, the model reaches its zenith in predictive performance, boasting an MAE of 0.132, MSE of 0.138, RMSE of 0.3715, and an outstanding R2 Score of 0.962.

The results indicate that the TFT-WQI-BP-TFL prediction model performs well in predicting the target variable. Higher epochs tend to yield better performance with lower MAE, MSE, and RMSE values, and higher R2 score, as depicted in Table LIII.

Table LIII. Performance of TFT-WQI-BP-TFL Model for Different Epochs

Dataset	Epochs	MAE	MSE	RMSE	R2 Score
WQI-BP	500	0.132	0.138	0.3715	0.962
	200	0.138	0.183	0.4278	0.957
	150	0.142	0.247	0.4970	0.932
	100	0.215	0.269	0.5187	0.913
	50	0.257	0.281	0.5301	0.89
	20	0.279	0.293	0.5413	0.88

Different dropout rates, specifically 0.2 and 0.3, were employed in a series of experiments aimed at constructing water quality index prediction models using the WQI-BP dataset and the knowledge gained from the TST-WQI-SA pre-trained model. The outcomes of these experiments, with respect to the evaluation metrics, for different dropout rates are detailed in Table LIV.

Table LIV. Prediction Results of TL Based WQI Prediction Models for Different Dropout Rates

Dataset	Models	Dropout	MAE	MSE	RMSE	R2 Score
WQI-BP	RNN-WQI-BP-TFL	0.3	0.235	0.246	0.4960	0.86
		0.2	0.371	0.387	0.6221	0.84
	LSTM-WQI-BP-TFL	0.3	0.195	0.137	0.3701	0.902
		0.2	0.229	0.163	0.4037	0.88
	GRU-WQI-BP-TFL	0.3	0.215	0.137	0.3701	0.892
		0.2	0.285	0.182	0.4266	0.87
	TFT-WQI-BP-TFL	0.3	0.132	0.138	0.3715	0.962
		0.2	0.142	0.247	0.4970	0.932

The prediction results of WQI models developed by through the TFT based pre-trained model and the WQI-BP dataset for various epochs and dropouts have been observed while implementing deep learning algorithms such as RNN, LSTM, GRU and the TFT architecture to discover the best prediction results. It is proved that the models trained with 500 epochs and dropout rate 0.3 with other hyperparameters such as adam optimizer, momentum as 0.8, and activation function as relu produced the best results and are shown in Table LV and depicted in Fig. 7.7.

Table LV. Prediction Results of Various TL Based WQI Prediction Models with TFT based Pre-trained Model

Pretrained Model	Models	MAE	MSE	RMSE	R2 Score
TFT-WQI-SA	**TFT-WQI-BP-TFL**	**0.132**	**0.138**	**0.3715**	**0.962**
	RNN-WQI-BP-TFL	0.235	0.246	0.4960	0.86
	LSTM-WQI-BP-TFL	0.195	0.137	0.3701	0.902
	GRU-WQI-BP-TFL	0.215	0.137	0.3701	0.892

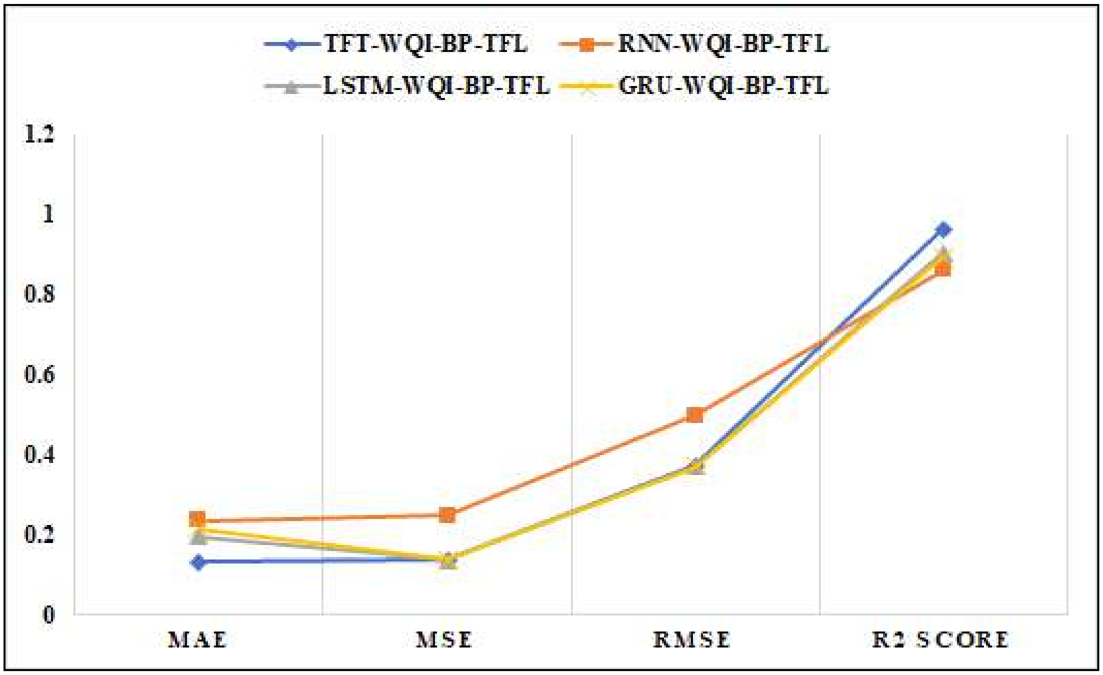

Fig.7.7. Prediction Performance of Various TL Based WQI Prediction Models with TFT based Pre-trained Model

Comparative Analysis of TL-based WQI Prediction Model Verus Base Models

The performance results of the TL based WQI prediction models built using the pre-trained model TFT-WQI-SA and with the WQI-BP dataset is compared with the prediction results of WQI models built with the Bharathapuzha river data. Among the models examined, the TFT-WQI-BP-TFL model stands out with a notably low MAE of 0.132, MSE of 0.138, RMSE of 0.3715, and an impressive R2 score of 0.962. Meanwhile, the RNN-WQI-BP-TFL model exhibits relatively higher MAE, MSE, and RMSE values of 0.235, 0.246, and 0.4960, respectively, yielding an R2 score of 0.86. The LSTM-WQI-BP-TFL model showcases competitive results, with an MAE of 0.195, MSE of 0.137, RMSE of 0.3701, and a robust R2 score of 0.902. Similarly, the GRU-WQI-BP-TFL model, with an MAE of 0.215, MSE of 0.137, RMSE of 0.3701, and an R2 Score of 0.892, demonstrates commendable predictive capabilities.

The performance results of prediction models built with and without transfer learning are tabulated in Table LVI and illustrated in Fig.7.8. From the prediction results it is proved that the prediction models built with transfer learning give better results. The TFT-WQI-BP-TFL model demonstrated superior performance results in predicting the water quality index, outperforming the other models in terms of MAE, MSE, RMSE, and R2 score. The inclusion of transfer learning appeared to contribute to the enhanced predictive capabilities of the TFT architecture. These results suggest that TFT-WQI-BP-TFL shows a promising choice for water quality index prediction.

Table LVI. Comparative Performance Analysis of TL-Based WQI Prediction Models Vs Base Models

Models	MAE	MSE	RMSE	R2 Score
TFT-WQI-BP-TFL	**0.132**	**0.138**	**0.3715**	**0.962**
RNN-WQI-BP-TFL	0.235	0.246	0.4960	0.86
LSTM-WQI-BP-TFL	0.195	0.137	0.3701	0.902
GRU-WQI-BP-TFL	0.215	0.137	0.3701	0.892
RNN-WQI-BP	0.598	0.523	0.7232	0.642
LSTM-WQI-BP	0.524	0.513	0.7162	0.684
GRU-WQI-BP	0.536	0.527	0.7259	0.672
TFT-WQI-BP	0.407	0.436	0.6603	0.705

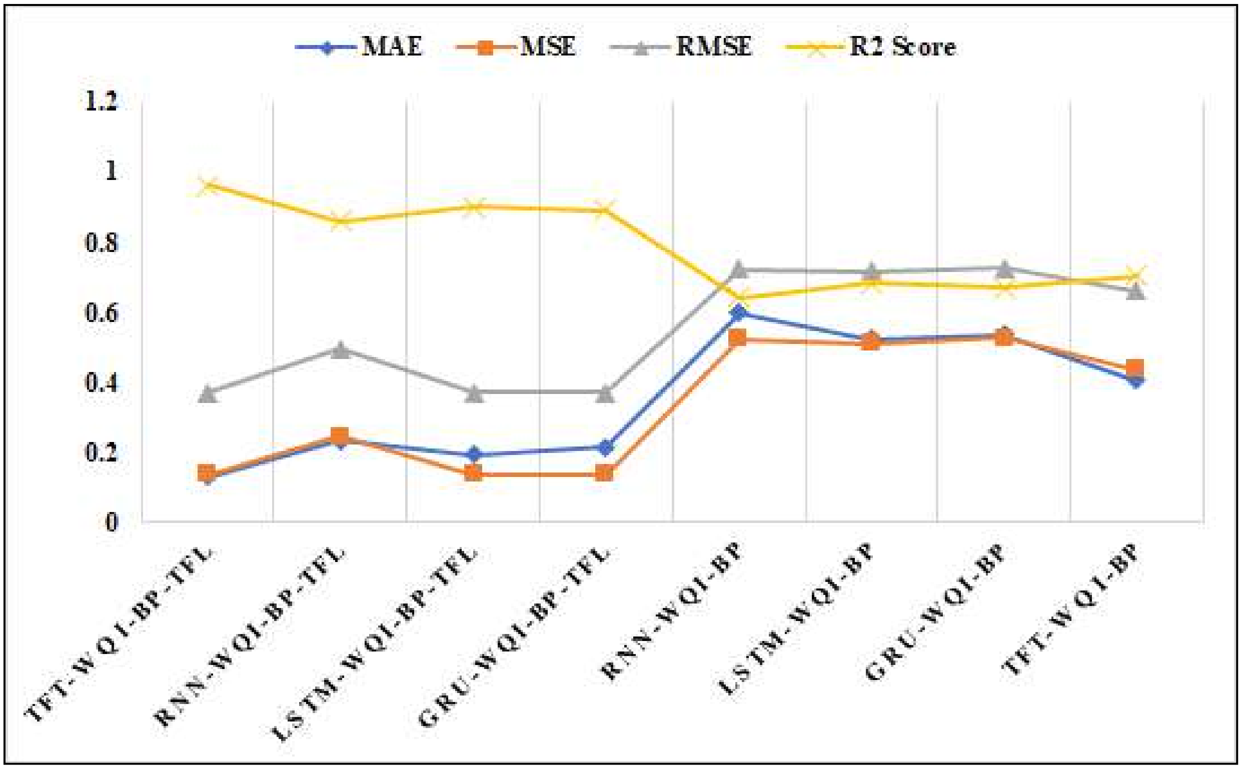

Fig.7.8. Performance Comparison of TL based WQI Prediction Models vs Base Models

Findings

The highlight of this work is the demonstration of effectiveness of temporal fusion transformers with transfer learning in predicting WQI. The prediction model benefits from the pre-trained knowledge, which helps to capture important features and patterns relevant to water quality. This transfer of knowledge enables the model to generalize better and make more accurate predictions. By utilizing the temporal fusion techniques, the model effectively fuses information from multiple time steps and generates robust predictions. The performance of the TL based WQI models demonstrates that the combined approach of transfer learning and TFT outperforms other WQI models described earlier. The integration of the transfer learning approach with the TFT can significantly enhance the prediction accuracy of the river WQI prediction model.

SUMMARY

A detailed walkthrough of building base WQI prediction models for the Bharathapuzha river, employing TFT and RNN variants, and showcasing the need for transfer learning was provided at the beginning of the chapter. The application of homogenous transfer learning approach in building enhanced river WQI prediction models was explored in this chapter in two segments. In the first segment, the LSTM based WQI model was used as a pre-trained model for the implementation of transfer learning with RNN, LSTM, and GRU using Bharathapuzha river data. In the second segment, the TFT based WQI model was used as a pre-trained model for the implementation of transfer learning with RNN, LSTM, GRU and also TFT using Bharathapuzha river data. The effect of homogenous transfer learning in building more efficient WQI models for the Bharathapuzha river was studied and presented. Additionally, it sets the stage for the upcoming chapter, which will expound upon the construction of an even more advanced WQI prediction model incorporating heterogeneous transfer learning.

Remarks

1. *TFT Architecture and RNN Variants for Water Quality Prediction of Bharathapuzha River - International Journal of Information Technology. (In Review)*
2. *Transfer Learning for Improved Deep Neural Network Models in River Water Quality Index Prediction - Journal of Network and Computer Application. (In Review)*
3. *Transfer Learning and Temporal Fusion Transformer for Enhanced River Water Quality Index Prediction - Expert System with Application. (In Review)*