9. CONCLUSION

The thesis titled ‘Temporal Fusion Transformer and Transfer Learning for Time Series River Water Quality Index Prediction Model’ demonstrated the research work carried out to predict river water quality. This research work showcases the development of novel prediction models for predicting water quality, utilizing cutting-edge deep learning techniques and transfer learning approaches.

Two distinct datasets have been generated using the physiochemical and seasonal parameters of Bhavani River water for the period January 2016 to December 2020, and two distinct datasets have been developed using water quality data of Bharathapuzha River for the period January 2020 to December 2021.

WQI prediction problem has been formulated as a regression task and WQI prediction models have been built using deep learning architectures such as RNN, LSTM, GRU and the specialized architecture TFT. The implementation is carried out in four phases. First, deep learning algorithms such as RNN, LSTM, and GRU were implemented using WQI-PCA and WQI-SA datasets to build accurate WQI prediction models. Secondly, the TFT was used to build improved WQI prediction model using the WQI-SA dataset. In the third phase, a homogenous transfer learning technique was applied with LSTM based pre-trained WQI model to train the networks RNN, LSTM, GRU independently with limited instances in WQI-BP dataset and an enhanced WQI prediction model has been built. The TFT-based pre-trained model jointly with the WQI-BP dataset was used with homogenous transfer learning and a hybrid WQI prediction model has been built. In the last phase, heterogeneous transfer learning was applied with the TFT based pre-trained model and an extended Bharathapuzha WQI-EBP dataset to build a robust WQI prediction model by training TFT.

Exhaustive experiments have been carried out using python libraries with Tensor flow, Keras and Scikit learn. The performances of various WQI prediction models have been evaluated with metrics such as mean absolute error, mean squared error, root mean squared error, and R2 score and the systematic investigation has been carried out. A thorough performance analysis of the WQI models is performed and the following observations are made.

- The prediction rate of WQI models is increased through representation learning of the time series sequence data with RNN, LSTM, and GRU networks.

- EDA performed on the raw time series data enabled to gain the insights of the data which facilitates data standardization.
- Through feature selection, the association between the pool of predictors and the target variable is strengthened which enables deep neural network architectures RNN, LSTM, and GRU to improve the learning of trends in the data.
- Promising results are produced by WQI prediction models as a result of augmenting seasonal data with regular physicochemical properties for training.
- The WQI-prediction model constructed using the TFT approach achieves encouraging results by leveraging the data handling capabilities of the transformer architecture and capturing the temporal patterns in the time series data.
- It is evident that LSTM and TFT-based WQI predictive models can be efficiently utilized as pre-trained models for transfer learning.
- Enhanced by transfer learning, WQI prediction models demonstrate improved accuracy even when trained on datasets with limited instances.
- The hybrid WQI prediction model developed using TFT and homogenous transfer learning has proved to be the good prediction model for forecasting WQI.
- The heterogeneous transfer learning approach implemented with the extended Bharathapuzha dataset and TFT based pretrained model has produced a robust generalized model for predicting WQI.

From the comprehensive experimental results, it is evident that the said objectives of this research work have been met. The research contributions made in this work are summarized below.

- Water quality prediction problem is formulated as a regression task and a suitable solution is provided using deep learning architectures and transfer learning approaches.
- Sufficient real-time data has been collected for learning the trends and patterns in water quality data and EDA has been performed to understand the statistical distribution of the data.
- Significant features are identified and captured from the data collected from the Bhavani River and Bharathapuzha River and four new time series datasets have been developed.
- Specialized deep learning architectures are used to design and develop the WQI prediction models by training time series data and to forecast water quality.
- TFT is used for constructing the WQI prediction model as TFT enhances the interpretability of time series forecasting by enduring temporal variations, and significant events.

- TFT is integrated with transfer learning approaches both homogeneous and heterogeneous and developed hybrid WQI prediction models by leveraging the merits of both TFT and transfer learning.
- A generalized and holistic framework has been developed using transfer learning to predict water quality index which can be applied to any river in India.

The potential challenge of data representativeness and domain shift can be revisited as a scope of future work and the research can be extended to develop a water quality monitoring system for predicting the river water quality of any geographical area by integrating the generalized WQI prediction model. The advancements in the Internet of Things, could enhance the development of predictive models by harnessing the capabilities of predictive analytics on devices or sensors, which could lead to further improvements in real-time monitoring system at any point of time.