ABSTRACT

Water is essential for a variety of purposes such as drinking, farming, and maintaining ecological balance. The quality of water relies on its chemical, physical, and biological characteristics. Challenges arise due to complex data sources, intertwined parameters, and climate changes, which complicate predictions. Existing research in water quality index prediction involves the integration of machine learning algorithms and environmental data to develop accurate models for assessing and forecasting water quality levels. But incorporating temporal patterns and diverse environmental parameters can lead to certain challenges in data collection, data modelling and efficient model building in the context of water quality index prediction. Recently, deep learning algorithms excel in the realm of building water quality index prediction models, by effectively capturing intricate patterns in complex environmental datasets, leading to enhanced predictive capabilities and more precise water quality prediction. Hence it is proposed to build a robust water quality index (WQI) prediction model using deep learning techniques by considering two river water sources - Bhavani River and Bharathapuzha river.

This research titled ‘Temporal Fusion Transformer and Transfer Learning for Time Series River Water Quality Prediction Model’ aims at developing a generalized river water quality index model using deep learning architectures and transfer learning techniques. The core objectives of this research are as follows.

- To design and develop an accurate WQI prediction model using physiochemical parameters and Recurrent Neural Networks (RNN) and its variants.
- To construct an improved water quality prediction model using physiochemical and seasonal parameters and RNN and its variants.
- To build an enhanced WQI prediction model using multi-horizon forecasting with a Temporal Fusion Transformer (TFT).
- To create an efficient WQI prediction model using the pre-trained model with RNN variants and homogenous transfer learning.
- To develop a hybrid time series WQI prediction model using temporal fusion transformer and homogenous transfer learning approach.
- To develop a robust time series WQI prediction model using temporal fusion transformer and heterogeneous transfer learning approach.

The problem of the river water quality index prediction is formulated as a regression task and the solution is proposed by training the significant features representing river water quality using deep learning architectures such as Recurrent Neural Network, Long Short- Term Memory (LSTM), Gated Recurrent Unit (GRU), the more sophisticated architecture Temporal Fusion Transformer and incorporating homogenous and heterogenous transfer learning.

Four independent time series datasets have been developed and used to build the river water quality index models. The first dataset is created by using the 26 physicochemical parameters collected across the eleven sampling stations of Bhavani River, for the period January 2016 to December 2020 with 10560 instances. It also contains 4 spatial parameters and the calculated WQI for each instance, and the dataset is named as WQI-PCA. The second dataset includes the 10 seasonal parameters collected from the visual crossing site which are pooled with 26 physicochemical parameters and 4 spatial attributes, the calculated WQI for each of the 10560 instance and the dataset is referred to WQI-SA dataset. The third dataset is developed by collecting data from the three monitoring stations of the Bharathapuzha River for the period January 2020 to December 2021 with 2190 instances. The dataset contains 26 physicochemical parameters and 10 seasonal parameters and 4 spatial parameters, the dataset is called WQI-BP. The fourth and final dataset developed for this study is the extended dataset of WQI-BP, where two more attributes such as SAR and flow rate are added with regular parameters of Bharathapuzha River data and is named as WQI-EBP.

The research work is carried out in four phases. In the first phase, deep neural network architectures such as RNN, LSTM and GRU are employed to build WQI prediction models as these architectures are significant in training sequence data. The WQI-PCA dataset is used to train the networks RNN and its variants LSTM and GRU independently. Various hyperparameters are properly defined for training and accurate WQI prediction models are developed. The next case emphasizes the influence of seasonal data on water quality prediction, and their contribution in building an improved WQI prediction model. The instances having both physiochemical and seasonal parameters of the WQI-SA dataset are given as input to the input layer of networks such as RNN, LSTM and GRU. The network training is done independently by properly setting the hyperparameters and the improved WQI prediction models are built.

In the second phase, the Temporal Fusion Transformer technique is applied to design and develop an enhanced WQI forecasting model. TFT is an innovative deep learning architecture adept at effectively modelling and forecasting time series data with complex temporal relationships. The WQI-SA dataset is used to train the TFT architecture with special hyperparameters and enhanced WQI prediction models are constructed.

In the third phase of this research, a transfer learning approach is adopted to boost the performance of WQI prediction models trained with limited data. This work uses the WQI-BP dataset and the models developed in previous phases as pre-trained models. The weights of the pre-trained WQI models are transferred while training RNN, LSTM, GRU with the WQI-BP dataset. Hyperparameters are properly configured and efficient WQI prediction models are built. The next case is to integrate two powerful techniques namely TFT and homogenous transfer learning and to develop a hybrid model to increase the performance of a WQI prediction model that is trained with limited data. This work uses the WQI-BP dataset, and the model developed in the previous phase as pre-trained model. The TFT based WQI prediction model is leveraged for informative insights, subsequently employed in the training RNN, LSTM, GRU with the WQI-BP dataset, facilitating the advancement of a novel WQI prediction model. While training and optimizing the model, the hyperparameters are established suitably, and upgraded WQI prediction models are built.

The fourth and final phase of the work incorporates heterogeneous transfer learning and develops an efficient WQI prediction model. This work uses the extended Bharathapuzha river data i.e., WQI-EBP dataset and the TFT based pre-trained model for training the TFT using heterogeneous transfer learning. Special hyperparameters utilized for TFT are properly configured and robust WQI prediction models are built.

Various experiments have been conducted using python libraries with TensorFlow, Keras, Scikit learn, and the performance of the various independent WQI prediction models are evaluated using metrics such as Mean Absolute Error, Mean Squared Error, Root Mean Squared Error, and R2 score with test dataset. The experimental results are investigated and the performance results of the WQI models are analysed. Finally, the research findings are summarized and reported.

Thus, this research enables the development of a generalized and holistic framework for building WQI prediction model using deep learning architectures and transfer learning approach to predict river water quality index. The integration of the WQI prediction models in real-time scenarios holds significant societal advantages by providing prompt and accurate water quality assessments. Ultimately, the implementation of the WQI prediction models fosters a sustainable environment and enhances the well-being of the broader society.