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

Incorporating Seasonal Trends for River Water Quality Prediction Models Using Deep Learning Algorithms

Jitha P. Nair[1*], Binu Mol T. V.[2], Deepika A.[3], Aparna Unnikrishnan[4], Muneer V. K.[5]

Submitted: 04/02/2024 **Revised**: 12/03/2024 **Accepted**: 18/03/2024

Abstract: In recent years, numerous contaminants have posed significant threats to rivers, streams, and lakes. The ability to analyse and predict water quality has become crucial in combating water pollution. Various seasonal factors, along with physicochemical properties, influence water quality over time. As water quality data forms a time series, the values of parameters fluctuate with changing meteorological conditions across seasons at each location. Consequently, robust time series analysis is essential for accurate water quality forecasting. Given the effectiveness of Recurrent Neural Networks for time sequence data, this study aims to develop a water quality prediction model by learning seasonal patterns in the time series dataset. The dataset comprises 10,560 unique instances that describe both physicochemical and seasonal factors. Predictive models are developed using RNN and its variants, Gated Recurrent Unit and Long Short-Term Memory and evaluated for their performance. The results demonstrate that incorporating seasonal data alongside regular physicochemical properties during model training significantly enhances predictive accuracy. By leveraging the temporal patterns inherent in the dataset, the models achieve promising results, indicating that the inclusion of seasonal variability is beneficial for improving water quality predictions. This approach not only highlights the importance of considering seasonal influences in water quality analysis but also showcases the potential of advanced neural network architectures in environmental monitoring and management. The study underscores the need for comprehensive data collection and sophisticated modelling techniques to effectively anticipate and mitigate the impacts of water contamination.

Keywords: *Deep Learning Architectures, Prediction Models, River Water Quality, Physicochemical Parameters, Seasonal Variations*

1. Introduction

The survival of the vast majority of living species, including humans, depends on water, making it the essential resource for life. High-quality water is crucial for all forms of life, as water-dependent species can only tolerate minimal pollution before their survival is threatened. When certain conditions are unmet, the survival of these species is jeopardized.

To effectively monitor and control water pollution, it is imperative to establish automatic water quality monitoring stations in key areas and develop accurate water quality prediction methods. Various water treatment techniques, such as anaerobic and aerobic treatments, activated sludge methods, and membrane bioreactor treatments, are employed to treat wastewater and mitigate pollution. Physicochemical treatment is utilized to separate colloidal particles in water.

Common physicochemical parameters used to determine the water quality index include conductivity, turbidity, total alkalinity, chloride, ammonia, hardness, sulfate, sodium, phosphate, boron, potassium, BOD, fluoride, nitrate, coliform, and dissolved oxygen. Water quality is also influenced by seasonal conditions, which vary throughout the year. In addition to physicochemical properties, seasonal attributes such as temperature, dew, humidity, precipitation, wind speed, and visibility are critical for predicting water quality.

Numerous research studies concentrate on a limited range of physicochemical parameters to develop water quality forecasting models. However, expanding the number of these factors and incorporating seasonal variables can substantially enhance the efficiency of water quality predictions.

Present computational methods in water quality prediction research encompass the grey relational method, mathematical statistics, model-based approaches, Bayesian approaches, genetic algorithms, MLP regressors, and support vector regressors. By broadening the range of parameters and integrating seasonal data, the accuracy and reliability of these predictive models can be significantly improved.

*1*Research Scholar, Department of Computer Science, PSGR Krishnammal College for Women, Peelamedu, Coimbatore, India*
2Assistant Professor, Department of Computer Science, KKTM Govt. College, Pullut, Thrissur, Kerala
3Assistant Professor, Department of Computer Science, Sri Ramakrishna College of Arts & Science, Coimbatore.
4Research Scholar, Department of Physics, PSGR Krishnammal College for Women, Peelamedu, Coimbatore
5Assistant Professor, Department of Computer Science, Sullamussalam Science College, Areekode, Kerala
**Corresponding Author: Jitha P Nair*
Jithapnair20@gmail.com

Chen et al. [1] developed a water quality prediction model using machine learning algorithms for the Huangpu River in Shanghai, China. They employed random forest and support vector machines (SVM) to predict water quality index (WQI) values, highlighting the importance of environmental factors and machine learning techniques for accurate predictions. Liang et al. [2] proposed a hybrid model combining neural networks and grey relational analysis for water quality prediction in the Han River, Korea. This approach integrated various water quality parameters to forecast trends and identify pollution sources, demonstrating effectiveness in improving prediction accuracy.

Zhang et al. [3] applied deep learning with convolutional neural networks (CNN) and long short-term memory (LSTM) networks to predict water quality in the Yangtze River, China. Their CNN-LSTM model captured spatial and temporal correlations, showing promise in enhancing prediction accuracy. Li et al. [4] studied ensemble learning techniques for water quality prediction in the Yellow River, China, comparing bagging, boosting, and stacking models. Ensemble methods combined multiple learners to achieve robust and accurate predictions of water quality parameters.

These studies advance water quality prediction by leveraging machine learning and deep learning techniques to address environmental challenges and improve prediction accuracy.

This research aims to develop an improved water quality prediction model by utilizing Recurrent Neural Networks (RNNs) to handle time series data. For this purpose, seasonal time series data collected from the Visual Crossing site, based on eleven sampling stations along the Bhavani River, is used. The water quality prediction model is built using variants of RNNs, specifically Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs). These models are then thoroughly evaluated to determine their effectiveness in predicting water quality.

2. Data Collection And Dataset Preparation

In our prior study, we developed a predictive model for water quality by examining trends in physicochemical features using time series data from river samples. The training dataset consisted of 26 physicochemical parameters, which included pH, conductivity, turbidity, phenolphthalein alkalinity, total alkalinity, chloride, COD, TKN, ammonia, calcium hardness, magnesium hardness, sulfate, sodium, TSS, TDS, FDS, phosphate, boron, potassium, BOD, fluoride, Nitrate-N, DO, TC, and FC, as detailed in Table 1.

Table 1: Sample Physicochemical Parameters Collected from Sampling Stations

pH	7.15	7.46	7.5	7.18	7.45	7.05	7.4	7.38	7.56	7.1
Conductivity	340	339	339	340	340	342	341	339	340	340
Turbidity	2	2	2	2	2	2	2	2	2	2
Phenolpth Alkalinity	0	0	0	0	0	0	0	0	0	0
Total Alkalinity	111	110	112	111	110	110	112	111	112	111
Chloride	21	21	22	21	20	20	20	21	21	21
COD	4	3.9	4	3.9	4	4	4	3.9	3.9	4
TKN	0.1	0.1	0.09	0.1	0.1	0.09	0.1	0.1	0.1	0.11
Ammonia	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
Hardness	118	118	119	119	119	119	118	118	118	117.5
Ca. hardness	74	74	74.5	74.5	74	73.5	73.5	73.5	74	74
Mg. Hardness	44	44	44	43.5	43.5	43.5	44	44	44	44
Sulphate	12	12.5	12	12	12.5	12.5	12	12	12.5	12
Sodium	27.1	27.1	27.2	27.2	27	27.1	27.1	27	27.1	27.1
TSS	300	300	300	300	300	300	300	300	300	300
TDS	190	190	189	189	189	190	189	190	189	188
FDS	174	174	174	174.5	174.5	174	174	174	173.5	173
Phosphate	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11
Boron	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Potassium	2.67	2.67	2.66	2.66	2.67	2.67	2.66	2.66	2.66	2.66
BOD	0.89	0.87	0.89	0.88	0.85	0.87	0.82	0.81	0.88	0.82
Fluoride	0.12	0.18	0.18	0.18	0.18	0.17	0.17	0.17	0.18	0.18
Nitrate-N	1.1	1.1	1.1	1	1.2	1	1.2	1.2	1.2	1.2
DO	6.99	6.97	6.81	7.19	7.3	7.39	7.06	7.02	6.97	7.39
TC	88	98	118	86	65	105	83	113	65	85
FC	80	80	80	79.5	79.5	79	79.5	80	80	80

Seasonal variations influence river water quality over time due to abrupt changes in climate. Some research articles indicates that these seasonal parameters significantly affect the Water Quality Index and its prediction in time series data. Rainfall and humidity are closely linked; relative humidity increases with rainwater evaporation. Wind speed is measured using a Davis Cup Anemometer at a height of three meters, compared to conventional measurements at ten meters. Higher wind speeds reduce the transition time between evaporative stages at low velocities. Dew, a crucial water source for rivers, significantly impacts microclimates and vegetation physiology. Global warming will alter precipitation distribution by changing air temperatures and circulation patterns. All seasonal factors alter the acceptable limits of physicochemical parameters, thereby reducing water quality. Consequently, this study considers seasonal features over the same time frame and incorporates their importance to enhance predictive model efficiency The seasonal features acquired from visual crossing sites are based on sampling station locations from January 2016 to December 2020 and the sample data is shown in Table 2. Seasonal Parameters like precipitation, precip over, cloud cover, humidity, dew, sea level pressure, wind speed, wind direction, and visibility are considered here, as these characteristics change with the season over time. These seasonal attributes are pooled with physicochemical parameters to develop a dataset in this work.

Table 2: Sample Seasonal Parameters Collected Visual Crossing Site

Temp	25	24	25	25	25	24	24	25	25	25
Dew	15.7	14.6	13.4	13.6	15.6	17.7	18.9	19.4	18.3	17.8
Humidity	59.3	56.72	51.89	53.06	58.8	62.79	68.91	68.63	65.71	63.8
Sea level pressure	1016.6	1017.1	1015.8	1015.7	1014.8	1014.8	1015.5	1015.5	1013.7	1014.5
Precipitation	0	0	0	0	0	0	0.2	0	0	0
Precip cover	0	0	0	0	0	0	4.17	0	0	0
Windspeed	16.3	14.4	13.1	15.4	14	18.7	40.2	13.6	14.4	14.9
Wind dir	52.9	62.3	61.7	68.2	56.5	69.3	114.6	95	94.9	65.1
Cloud cover	27.4	17.9	5.5	14.1	14.6	16	32.3	42.5	26.3	14
Visibility	5.5	6	5.7	5.9	5.6	5.5	4.8	5.3	5.1	5.4

The Water Quality Index (WQI) is a tool used to measure the quality of water. It is composed of several seasonal attributes that are used to assess the overall health of a water body. Temperature is an important attribute as it helps to indicate the presence of certain species, as well as the activity of the water body. Dew is a measure of the amount of water vapour present in the atmosphere. It is significant to determine water quality because it helps to regulate the temperature of the environment, and it can also indicate the amount of precipitation that is likely to occur. Humidity is a measure of the amount of water vapour present in the air. It is also important to measure the water quality as it can affect the rate of evaporation, and also indicate the temperature of the environment.

Sea level pressure is a measure of the atmospheric pressure at sea level and is important in water quality prediction because it can affect the rate at which water evaporates, and also indicate the amount of precipitation that is likely to occur. Precipitation is a measure of the amount of liquid or solid water particles that have fallen from the atmosphere. It is crucial for maintaining water quality as it can affect the amount of dissolved oxygen in the water, and it can also indicate the pollutants that are present. Precip Over is a measure of the amount of precipitation that has fallen over a certain period. It has a significant impact on the quality of water because it can indicate the number of pollutants that are present in the water, and it can also indicate the number of nutrients that are available for plant growth.

Wind speed is a measure of the speed of the wind that is blowing. It is important to water quality prediction because it can affect the rate at which water evaporates, and it can also indicate the pollutants that are present in the water. Wind direction is a measure of the direction in which the wind is blowing. It plays a vital role in water quality because it can affect the rate at which water evaporates, and it can also indicate the number of pollutants that are present in the water. Cloud cover is a measure of the amount of water vapour that is present in the atmosphere. It has a significant effect on water quality due to its ability to affect the rate of evaporation, and it can also indicate the volume of pollutants that are present in the water. Visibility is a measure of how far one can see in the atmosphere. It is a key factor in the quality of water because it can indicate the pollutants that are present in the water, and it can also indicate the number of nutrients that are available for plant growth.

Table 3: Water Quality Parameters

Physicochemical Parameters		**Seasonal Parameters**
pH	TSS	Temperature
Conductivity	TDS	Dew
Turbidity	FDS	Humidity
Phenolphthalein Alkalinity	Phosphate	Sea level pressure
Total Alkalinity	Boron	Precipitation
Chloride	Potassium	Precip cover
COD	BOD	Windspeed
TKN	Fluoride	Wind dir
Ammonia	Nitrate-N	Cloud cover
Hardness	TC	Visibility
Ca. hardness	FC	**Spatial Parameters**
Mg. hardness	Dissolved Oxygen	Station ID
Sulphate	**Temporal Parameter**	Latitude
Sodium	Date	Longitude

Thus, twenty-six physiochemical attributes are pooled with ten seasonal attributes along with spatial parameters to develop the WQI-SA dataset. Finally, there is a total of 40 attributes forming the time series data prepared for this research work.

The river water quality data undergoes exploratory data analysis to understand the data properties and assess the importance of each parameter in generating the water quality index. Physicochemical data from sampling stations and seasonal data from the Visual Crossing site are listed in Table 3. Statistical techniques such as heatmaps, boxplot analysis, pair plot analysis, and histograms are used to analyse and interpret the distribution of parameter values. Boxplot analysis reveals that seasonal parameters such as wind speed and cloud cover exhibit a wide range of values. While wind speed ranges between 10 and 270, and cloud cover is between 0 and 100. Therefore, the parameter values are normalised so that they lie within the usual range for each parameter. Wind speed and cloud cover are standardised using the min-max approach. Using Pearson correlation, the heatmap is used to visualise and analyse the correlation between the parameters, such as positive and negative. The bar graph analysis of wind speed, humidity, visibility, cloud cover, and physicochemical parameters are depicted in Figure 1a. pH, turbidity, cloud cover, FDS, boron, TC, TSS, and wind speed are the parameters that have a negative correlation with WQI and are displayed in Fig. 1b.

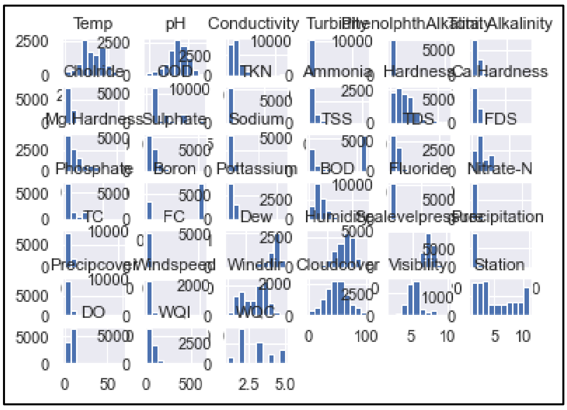

Fig.1a Bar graph analysis

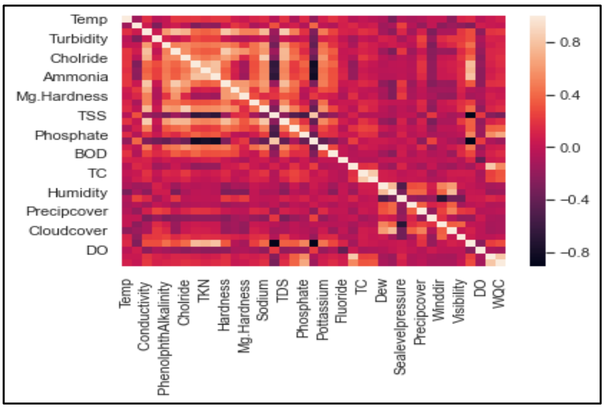

Fig.1b Heatmap analysis

EDA reveals that some instances in the dataset include missing values that must be eliminated, so data cleaning is performed. EDA explains better about the attribute distributions and parameter correlations, providing suitable solutions for data modelling and pre-processing needs.

The Water Quality Index (WQI) is a measure of the overall water quality of the proposed system. The WQI can be used to monitor changes in water quality over time and to assess the suitability of the water body. It is calculated by taking the average of several factors that are indicators of water quality, like temperature, pH, bod, and cod. The WQI is then assigned a score based on a range of 0 to 120, with higher scores indicating poor water quality. The WQI is computed and then added as the target variable along with the 40 independent variables for the WQI modelling prediction task. Hence in the work, the dataset includes both physiochemical and seasonal parameters and it contains 10560 instances. Feature selection is a vital phase in predictive modelling in which appropriate parameters that contribute significantly to predicting the target variable are chosen. In this study, the SelectKBest algorithm was utilized to identify important features for calculating the water quality index. According to the SelectKBest feature selection algorithm, conductivity ranked highest in estimating the water quality index, followed by ammonia and phosphate. Conversely, boron and phenolphthalein alkalinity, identified as less significant by the feature selection process, were removed from the dataset.

This feature selection method improved the river water quality dataset and finally the dataset with 10560 instances and 38 attributes has been developed and is named as WQI-SA dataset for reference.

3. Water Quality Index Prediction Model

The challenge of predicting the water quality index is framed as a regression problem and addressed using deep neural network architectures. Deep neural networks effectively characterize and classify data by processing input data along with associated weights and biases through multiple interconnected layers. These networks typically feature numerous layers, with visible input and output layers that enhance prediction accuracy. Figure 2 illustrates the architecture of the proposed framework for the WQI prediction model. In this framework, pre-processed data is fed into the deep learning model at the input layer, and predictions are generated at the output layer. The model benefits from large-scale data sets, with performance improving as more data is incorporated, resulting in high-quality predictions.

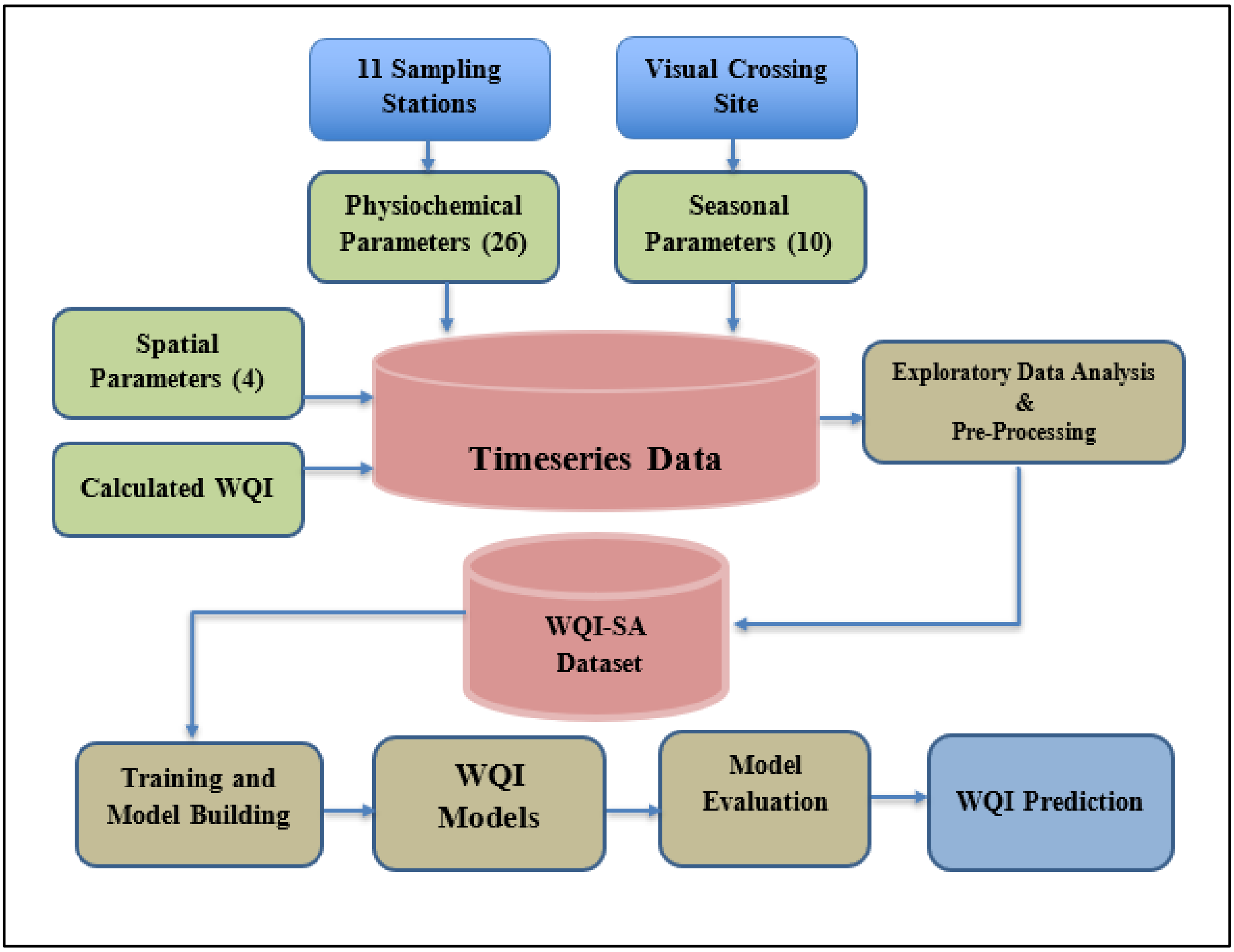

Fig. 2. Proposed WQI Model Architecture for River Water Quality Index Prediction

In this study, deep learning architectures such as recurrent neural networks (RNNs), long short-term memory (LSTM), and gated recurrent units (GRU) were selected to develop a river water quality index prediction model capable of handling sequence data. Recurrent Neural Networks (RNNs) use outputs from previous sections as inputs for subsequent sections, with the hidden state being crucial for storing sequence information. However, traditional RNNs are prone to the vanishing gradient problem due to their limited ability to retain long-term memory over many time steps. To address this, LSTM and GRU architectures were employed.

LSTM units are designed to remember and forget information as needed, maintaining an internal cell state vector to retain pertinent information from previous time steps. On the other hand, GRU units utilize update and reset gates to control the flow of information through the network, allowing for better gradient flow and improved long-term memory retention compared to standard RNNs.

In this research, 80% of the WQI-SA dataset instances were used to independently train RNN, LSTM, and GRU models, with hyperparameters optimized to improve model performance. Hyperparameters such as hidden layers, dense layers, optimizer type, epochs, momentum, batch size, activation functions, and dropout rates were fine-tuned to enhance model accuracy and predictive capability.

Hidden layers are positioned between input and output layers, and dense layers connect all neurons from the previous layer to the next, significantly improving accuracy. Optimizers adjust neural network parameters like weights and learning rates to minimize loss and enhance model performance. Epochs determine the number of complete dataset iterations during training, while momentum enhances gradient-based optimization by considering gradients from previous steps. Activation functions introduce nonlinearity, crucial for separating and reducing output dimensions from dense layers.

Lastly, dropout layers were used to prevent overfitting by randomly excluding specific layer weights during training, ensuring the model's generalization ability. This study successfully constructed WQI prediction models using LSTM, GRU, and RNN architectures, employing proper hyperparameter settings to facilitate representation learning from input instances.

The learning rate determines the speed at which a deep model replaces a previously learned concept with a new one. Finally, three independent WQI prediction models are built by learning water quality patterns from the input instances of the WQI-SA dataset through training RNN, LSTM and GRU with proper hyperparameters settings. These models are called as RNN-WQI-SA, LSTM-WQI-SA and GRU-WQI-SA models for reference. The effectiveness of the WQI forecasting models is evaluated using MAE, MSE, RMSE and R2 score.

4. Experiment And Results

In our previous study, experiments were conducted using the time series WQI-PCA dataset, which includes samples with physicochemical parameters. Prediction models were developed using deep neural architectures, specifically RNN, GRU and LSTM. The prediction results of these models are summarized in Table 4, revealing that the GRU-based WQI prediction model achieved an accuracy of 84% in predicting WQI.

Table 4: Prediction Result using WQI-PCA Dataset

Dataset	Model	MAE	MSE	RMSE	R2 Score
WQI-PCA	RNN-WQI-PCA	0.512	0.408	0.6387	0.8
	LSTM-WQI-PCA	0.393	0.2401	0.4900	0.838
	GRU-WQI-PCA	0.364	0.2098	0.4580	0.845

In the study, deep learning algorithms such as GRU, LSTM and RNN were employed to train the WQI-SA dataset from the Bhavani River using Python libraries. The dataset consisted of 8124 tagged instances for training, and evaluation of the prediction models was conducted using metrics such as MAE, MSE, RMSE, and R2 score values. The test dataset comprised 2009 tagged instances from the WQI-SA dataset.

For the deep learning models, hyperparameters were specified as follows: the dense layer units ranged from 5 to 10, and the Adam optimizer was used. Epoch sizes of 20, 50, 100, 150, 200, and 500 were experimented with. The ReLU activation function was chosen for training, and momentum was varied between 0.5 and 0.9. Initially, a dropout rate of 0.2 was considered, but later, 0.3 was found to yield better results. The learning rate was set at 0.1, and the batch size was alternated between 32 and 64. Experimental results indicated that setting the momentum to 0.8, using an epoch size of 500, dropout rate of 0.3, and ReLU activation function produced the most accurate predictions.

These settings were selected through rigorous experimentation to optimize the performance of deep learning models for predicting water quality index. The results of the RNN-based WQI prediction model (RNN-WQI-SA model) are experimented with various epochs such as from 20 to 500 where various metrics are measured at different epochs. At epoch 500, the RNN model achieves an MAE of 0.424, indicating the average absolute difference between the predicted and actual values. The MSE is calculated as 0.384, representing the average of squared differences. The RMSE is 0.6196, which is the square root of the MSE. The R2 score, measuring the goodness of fit, is 0.82, indicating a high level of prediction accuracy. Moving to epoch 200, the MAE increases slightly to 0.459, while the MSE becomes 0.392. The RMSE is 0.6260, and the R2 score remains relatively high at 0.813. As the number of epochs decreases, the MAE and MSE values gradually increase, indicating a larger difference between the predicted and actual values.

At epoch 150, the MAE is 0.482, and the MSE is 0.424, resulting in an RMSE of 0.6511. The R2 score decreases to 0.806, suggesting a slightly lower level of prediction accuracy compared to the previous epochs. At epoch 100, the MAE increases further to 0.512, and the MSE becomes 0.462. The RMSE is 0.6797, and the R2 score remains relatively stable at 0.80. With only 50 epochs, the MAE reaches 0.537, and the MSE increases to 0.527. The RMSE becomes 0.7259, while the R2 score decreases slightly to 0.79. Finally, at epoch 20, the MAE is 0.579, the MSE is 0.561, and the RMSE is 0.7489. The R2 score drops to 0.78. These values reflect the performance of the RN-WQI-SA model on the WQI-SA dataset at different epochs, providing insight into the prediction results which are tabulated in Table 5.

Table 5. Evaluation of RNN-WQI-SA Model Performance Using Different Epoch

Dataset	Epochs	MAE	MSE	RMSE	R2 Score
WQI-SA	20	0.579	0.561	0.7489	0.78
	50	0.537	0.527	0.7259	0.79
	100	0.512	0.462	0.6797	0.8
	150	0.482	0.424	0.6511	0.806
	200	0.459	0.392	0.626	0.813
	500	0.428	0.384	0.6196	0.82

The prediction results of the LSTM-based WQI prediction model (LSTM-WQI-SA model) for different epochs on the WQI-SA dataset. At epoch 500, the LSTM-WQI-SA model achieves an MAE of 0.298, indicating the average absolute difference between the predicted and actual values. The MSE is calculated as 0.2084, representing the average of squared differences. The RMSE is 0.4565, which is the square root of the MSE. The R2 score, measuring the goodness of fit, is 0.856, indicating a high level of prediction accuracy. Moving to epoch 200, the MAE increases slightly to 0.304, while the MSE becomes 0.239. The RMSE is 0.4888, and the R2 score remains relatively high at 0.85. As the number of epochs decreases, the MAE and MSE values gradually increase, indicating a larger difference between the predicted and actual values. At epoch 150, the MAE is 0.328, and the MSE is 0.274, resulting in an RMSE of 0.5234. The R2 score decreases to 0.843, suggesting a slightly lower level of prediction accuracy compared to the previous epochs.

At epoch 100, the MAE increases further to 0.371, and the MSE becomes 0.291. The RMSE is 0.5394, and the R2 score remains relatively stable at 0.839. With only 50 epochs, the MAE reaches 0.398, and the MSE increases to 0.328. The RMSE becomes 0.5727, while the R2 score decreases slightly to 0.83. Finally, at epoch 20, the MAE is 0.402, the MSE is 0.367, and the RMSE is 0.6058. The R2 score drops to 0.827. These values illustrate the performance results of the LSTM-WQI-SA model on the WQI-SA dataset at different epochs, providing insight into the prediction results which are tabulated in Table 6.

Table 6. Evaluation of LSTM-WQI-SA Model Performance Using Different Epoch

Dataset	Epochs	MAE	MSE	RMSE	R2 Score
WQI-SA	20	0.402	0.367	0.6058	0.827
	50	0.398	0.328	0.5727	0.83
	100	0.371	0.291	0.5394	0.839
	150	0.328	0.274	0.5234	0.843
	200	0.304	0.239	0.4888	0.85
	500	0.298	0.2084	0.4565	0.856

The prediction results of the GRU-based WQI prediction model (GRU-WQI-SA model) for different epochs on the WQI-SA dataset. At epoch 500, the GRU-WQI-SA model achieves an MAE of 0.39, indicating the average absolute difference between the predicted and actual values. The MSE is calculated as 0.2149, representing the average of squared differences. The RMSE is 0.4636, which is the square root of the MSE. The R2 score, measuring the goodness of fit, is 0.839, indicating a relatively high level of prediction accuracy. Moving to epoch 200, the MAE increases slightly to 0.412, while the MSE becomes 0.2342. The RMSE is 0.4839, and the R2 score decreases to 0.83. As the number of epochs decreases, the MAE and MSE values gradually increase, indicating a larger difference between the predicted and actual values.

At epoch 150, the MAE is 0.436, and the MSE is 0.269, resulting in an RMSE of 0.5187. The R2 score decreases to 0.823, suggesting a slightly lower level of prediction accuracy compared to the previous epochs. At epoch 100, the MAE increases further to 0.452, and the MSE becomes 0.287. The RMSE is 0.5357, and the R2 score remains relatively stable at 0.82. With only 50 epochs, the MAE reaches 0.462, and the MSE increases to 0.315. The RMSE becomes 0.5612, while the R2 score decreases slightly to 0.803. Finally, at epoch 20, the MAE is 0.474, the MSE is 0.348, and the RMSE is 0.5899. The R2 score drops to 0.793. These values highlight the performance of the GRU model on the

WQI-SA dataset at different epochs, providing insight into the prediction results which are tabulated in Table 7.

Table 7. Evaluation of GRU-WQI-SA Model Performance Using Different Epoch

Dataset	Epochs	MAE	MSE	RMSE	R2 Score
WQI-SA	20	0.474	0.348	0.5899	0.793
	50	0.462	0.315	0.5612	0.803
	100	0.452	0.287	0.5357	0.82
	150	0.436	0.269	0.5187	0.823
	200	0.412	0.2342	0.4839	0.83
	500	0.39	0.2149	0.4636	0.839

Various experiments have been carried out with different dropout rates such as 0.2 and 0.3 for building WQI prediction models using the WQI-SA dataset and the experimental results concerning the same evaluation metrics are shown in Table 8.

Table 8. Results of WQI Prediction Models for Different Dropout Rates

Dataset	Algorithm	Dropout	MAE	MSE	RMSE	R2 Score
WQI-SA	RNN	0.2	0.428	0.384	0.6197	0.82
		0.3	0.482	0.424	0.6512	0.806
	LSTM	0.2	0.298	0.2084	0.4565	0.856
		0.3	0.328	0.274	0.5235	0.843
	GRU	0.2	0.39	0.2149	0.4636	0.839
		0.3	0.436	0.269	0.5187	0.823

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 and dropout rate 0.3 with other hyperparameters like 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 9 and depicted in Fig. 3.

Table 9. Performance Analysis of Three WQI Prediction Models

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

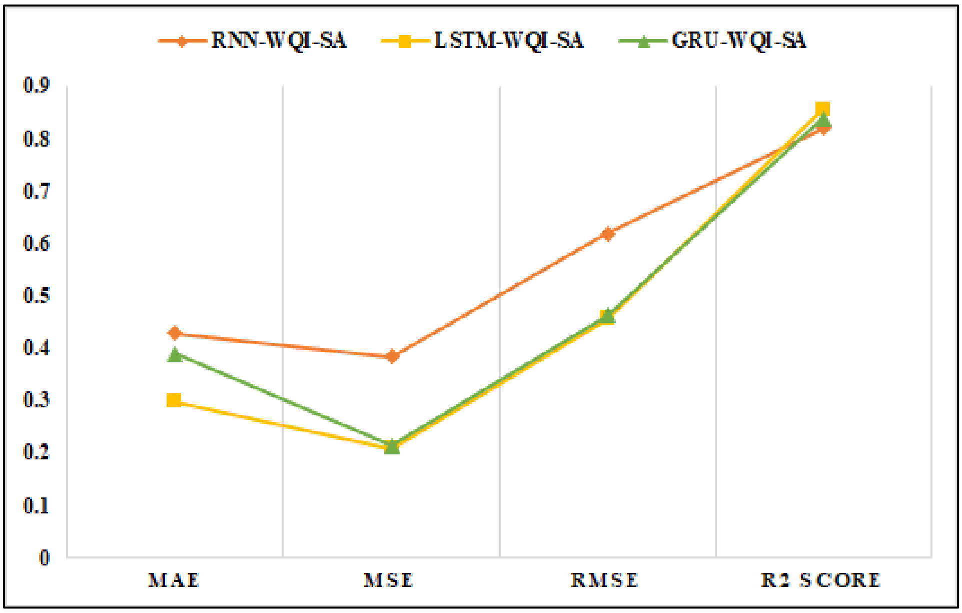

Fig.3. Prediction Performance of all Three WQI Models

Based on the above results, it is evident that the LSTM-based WQI prediction model demonstrates promising performance, achieving a high R2 score and lower error rates. Specifically, the mean absolute error of the LSTM-based forecasting model is lower compared to the RNN and GRU algorithms. Moreover, the root mean squared error is also lower for the LSTM-WQI-SA model compared to the RNN-WQI-SA and GRU-WQI-SA models. The higher R2 score of the LSTM-WQI-SA forecasting model indicates greater accuracy compared to the other prediction models.

Comparative Analysis WQI Models based on WQI-PCA and WQI-SA Datasets

The performance results of prediction models built using two distinct datasets such as WQI-PCA and WQI-SA are compared to analyse the efficiency of the prediction models. For the WQI-PCA dataset, the RNN-WQI-PCA model achieved an MAE of 0.512, MSE of 0.408, RMSE of 0.6387, and an R2 Score of 0.8. The LSTM-WQI-PCA model outperformed RNN-WQI-PCA with an MAE of 0.393, MSE of 0.2401, RMSE of 0.49, and an R2 Score of 0.838. The GRU-WQI-PCA model showed the best performance on the WQI-PCA dataset with an MAE of 0.364, MSE of 0.2098, RMSE of 0.4580, and an impressive R2 Score of 0.845. It is evident that the GRU-WQI-PCA model yielded the most accurate predictions among the models evaluated.

The prediction results of the models built using the WQI-SA dataset show promising results as compared to previous work. The RNN-WQI-SA model demonstrated an MAE of 0.428, MSE of 0.384, RMSE of 0.6197, and an R2 Score of 0.82. The LSTM-WQI-SA model performed even better, achieving an MAE of 0.298, MSE of 0.2084, RMSE of 0.4565, and an R2 Score of 0.856, indicating high predictive accuracy. The GRU-WQI-SA model also delivered good results with an MAE of 0.39, MSE of 0.2149, RMSE of 0.4636, and an R2 score of 0.839. The results clearly indicate that the LSTM-WQI-SA model produced the most precise predictions among the other prediction models.

From the comparative study, it is evident that the prediction models built using the WQI-SA dataset performed better than the models built using the WQI-PCA dataset. The LSTM-WQI-SA model emerged as the most accurate one, exhibiting the lowest MAE, MSE, and RMSE, along with the highest R2 Score. Here it is evident that the incorporation of seasonal parameters has improved the efficacy of the prediction models. The performance analysis of the WQI prediction models is tabulated in Table 10 and illustrated in Fig. 4.

Table 10. Performance Comparison of WQI Models based on WQI-PCA and WQI-SA Datasets

Dataset	Models	MAE	MSE	RMSE	R2 Score
WQI-PCA	RNN-WQI-PCA	0.512	0.408	0.6387	0.8
	LSTM-WQI-PCA	0.393	0.2401	0.4900	0.838
	GRU-WQI-PCA	0.364	0.2098	0.4580	0.845
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

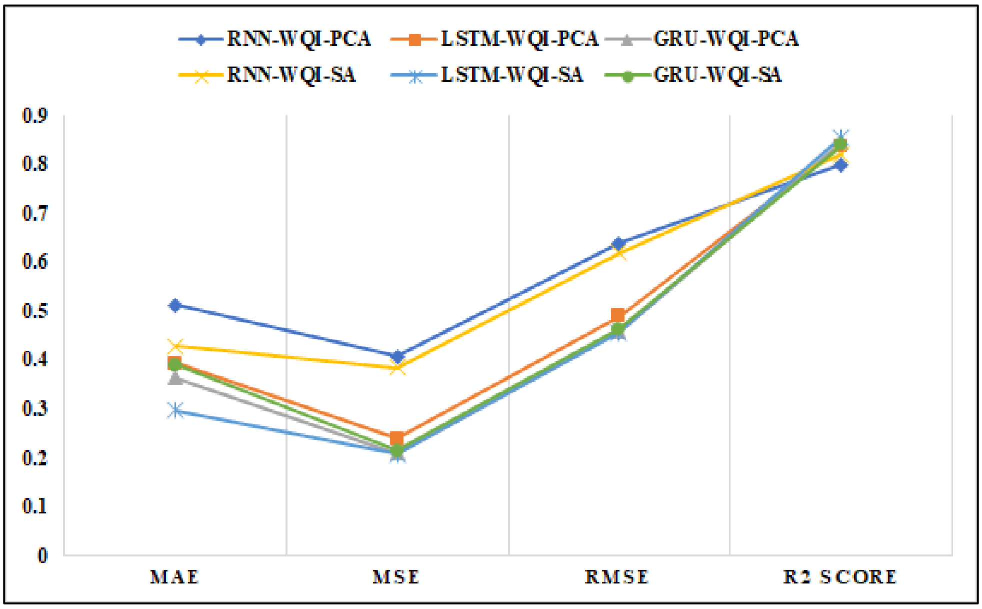

Fig. 4. Performance Comparison of WQI Models based on WQI-PCA and WQI-SA Datasets

Findings

This study demonstrates the efficacy of deep learning approaches in developing predictive models for WQI using time series data, particularly by incorporating seasonal parameters. Seasonal parameters significantly enhance WQI prediction by strengthening the relationship between predictors and the target variable, thereby aiding LSTM, RNN, and GRU networks in learning data trends more effectively. These models benefit from the self-extracted features learned within the networks, leading to improved prediction rates. Proper hyperparameter configuration during training further reduces error rates, making the enhanced water quality prediction model with seasonal time series data a robust tool for accurately predicting water quality.

5. Conclusion

This study underscored the significance of seasonal data in constructing Water Quality Index (WQI) prediction models. Deep learning architectures were applied to river water quality time series forecasting, demonstrating their efficacy in achieving accurate WQI predictions. Seasonal data collected from the Visual Crossing site between 2016 and 2020 was combined with physicochemical parameters from the Bhavani River to create a new time series dataset. The river water quality forecasting model was designed and implemented using deep learning architectures such as LSTM, RNN, and GRU. The performance of these models was evaluated and compared with models trained solely on physicochemical parameters. The evaluation results indicated that incorporating seasonal data significantly improved the efficiency of the water quality prediction model. A generalized model was developed, capable of predicting the water quality of any river. Furthermore, the developed model can serve as a pre-trained model for transfer learning applications.

References

[1] Chen, Xiang, et al. "Machine Learning-Based Water Quality Prediction for the Huangpu River in Shanghai, China." Environmental Monitoring and Assessment, vol. 193, no. 5, 2021, article 277.

[2] Liang, Changwei, et al. "Hybrid Neural Network and Grey Relational Analysis Model for Water Quality Prediction in the Han River, Korea." Journal of Hydrology, vol. 589, 2020, article 125053.

[3] Zhang, Xiaojun, et al. "Deep Learning for Water Quality Prediction Using Convolutional and Long Short-Term Memory Neural Networks: A Case Study of the Yangtze River, China." *Water Research*, vol. 184, 2020, article 116197.

[4] Li, Yuyang, et al. "Ensemble Learning-Based Water Quality Prediction in the Yellow River, China." *Science of the Total Environment*, vol. 716, 2020, article 137041.

[5] These references provide the necessary details for each article in MLA format.Liu J, Yu C, Hu Z, Zhao Y, Bai Y, Xie M, Luo J (2020) Accurate prediction scheme of water quality in smart mariculture with deep Bi-S-SRU learning network. IEEE Access 8:24784–24798

[6] Yahya A, Saeed A, Ahmed AN, Binti Othman F, Ibrahim RK, Afan HA, Elshafie A (2019) Water quality prediction model-based support vector machine model for ungauged river catchment under dual scenarios. Water 11(6):1231

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

[8] Nair, Jitha P., and M. S. Vijaya. ‘River Water Quality Prediction and Index Classification Using Machine Learning’. *Journal of Physics: Conference Series*, vol. 2325, no. 1, Aug. 2022, p. 012011.

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

[10] Nair, J.P., Vijaya, M.S. (2023). Exploratory Data Analysis of Bhavani River Water Quality Index Data. In: Kumar, S., Hiranwal, S., Purohit, S.D., Prasad, M. (eds) Proceedings of International Conference on Communication and Computational Technologies. Algorithms for Intelligent Systems. Springer, Singapore.

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

[12] Li, G., 2006. Stream temperature and dissolved oxygen modelling in the Lower Flint River Basin. PhD Dissertation. University of Georgia, Athens, GA.

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

[14] Li L, Jiang P, Xu H, Lin G, Guo D, Wu H (2019) Water quality prediction based on recurrent neural network and improved evidence theory: a case

study of Qiantang River, China. Environ Sci Pollut Res 26(19): 19879–19896

[15] Nair, Jitha & Vijaya, M.S (2023). Design And Development of Efficient Water Quality Prediction Models Using Variants Of Recurrent Neural Networks. European Chemical Bulletin. 12. 1210 – 1223. 10.31838/ecb/2023.12.si5.0143.

[16] Roy, Retsy Ann, Jitha P. Nair, and Elizabeth Sherly. "Decision tree based data classification for marine wireless communication." 2015 International Conference on Computing and Network Communications (CoCoNet). IEEE, 2015.

[17] G Tan, J Yan, C Gao, and S Yang, Prediction of water quality time series data based on least squares support vector machine, Procedia Engineering, Vol. 31, 2012, pp. 1194-1199.

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

[19] Aldhyani, T. H. H., Al-Yaari, M., Alkahtani, H. & Maashi, M. Water quality prediction using artificial intelligence algorithms. Applied Bionics and Biomechanics 2020, 6659314 (2020).

[20] Nair, Jitha P., and M. S. Vijaya. "Temporal fusion transformer: A deep learning approach for modelling and forecasting river water quality index." International Journal of Intelligent Systems and Applications in Engineering 11.10s (2023): 277-293.