WATER QUALITY MONITORING SYSTEM USING RECURRENT NEURAL NETWORK AND MULTILAYER PERCEPTION

[1]ABDULRAUF TOSHO, [2]RAJI BOLA, [3]SAKA KAMIL, AZEEZ DIEKOLA, TAJUDDEN KAFAYAT, ADEBAYO AMINAT
[123456]Department of Computer Science, Al-Himah University Ilorin, Nigeria
Abdtosh@gmail.com[1] lanbogold@gmail.com[2], kamilsaka67@gmail.com[3], diekola025@gmail[4].com,kotajudeen@alhikmah.edu.ng[5], & ayakub@alhikmah.edu.ng[6]

Abstract
Monitoring water quality is crucial to sustaining healthy aquatic habitats and maintaining safe drinking water. Conventional water quality monitoring systems frequently depend on manual testing and sampling, which can be labor-intensive, time-consuming, and prone to delayed responses. This paper developed an intelligent water quality monitoring system using Recurrent Neural Networks (RNN) and Multilayer Perceptions (MLP) to automate the analysis of water quality data and provide real-time predictions. Recurrent Neural Networks (RNNs) was used as feature extraction algorithm and Multilayer perception (MP) was used to classified the feature extracted from brisbane water quality dataset. The Recurrent Neural Network model was evaluated for its reliability in predicting temperature, hydrogen, salinity, and specific conductance. To assess the effectiveness of the regression models, three assessment metrics were computed: Mean Absolute Error (MAE), Mean Square Error (MSE), and coefficient of determination (R2). The model had a mean square error of 1.33 and an average absolute difference of 0.81, indicating reasonable reliability. In terms of classification, the testing findings showed that the Multilayer Perception model produced the good results, with an accuracy of 98% when predicting WQC values. The RNN model is recommended for water quality prediction due to its superior efficiency and accuracy, especially in pH indicators.

Keywords: Water Quality Control System, Recurrent Neural Networks, Brisbane Dataset, Multilayer Perception

Introduction

Nowadays the quality of water is depleting day by day, this causes various diseases and deaths (Sana *et al.,* 2022). The surface water like rivers, lakes, ponds can be easily contaminated by throwing human waste, industrial waste and waste pollution (Theofanis, Christos & Marios, 2014). Water Quality Management system (WQYs) needs fast and accurate system that gives high accuracy. The existing system is not sufficient to give high accuracy (Smith & Brown, 2019). The old process of monitoring water quality like manually collecting the water and tests in labs which are very costly and ineffective process (Nikhil Kumar Koditala, 2018).

Water distribution systems are indispensable infrastructure elements, crucial for sustaining urban life and ensuring the safety and well-being of populations (Tsotsope, *et al.,* 2020). These systems facilitate the delivery of clean and safe water to both residential and commercial users, presenting numerous operational challenges including maintenance, efficiency, and demand management (Moonsun *eta al.,* 2022). The efficient management of these systems becomes increasingly complex with the expanding scope of urbanization and the unpredictable shifts in environmental conditions (Liu *et al.,* 2022). Traditional methodologies in water management often characterized by manual oversight and basic automation are increasingly inadequate for the demands of modern urban environments. These methods are typically reactive rather than proactive, and struggle to scale or adapt to dynamic conditions until Internet of Things implemented (Smith & Brown, 2019).

This research on developing an efficient Recurrent Neural Network (RNN) model for water monitoring in IoT environments holds substantial significance across multiple dimensions of urban infrastructure management and technological innovation. The RNN model enables water authorities to optimize resource allocation, reducing both overuse and shortages. This can lead to significant cost savings and more sustainable water usage patterns

This research paper focused on developing a highly efficient RNN model, meticulously optimized for realistic IoT-enabled environments within water quality monitoring frameworks. The model is designed to integrate seamlessly with IoT architectures, processing data from a myriad of sensors to optimize water quality, enhance predictive maintenance, and improve system reliability. This study not only addresses the immediate operational efficiencies but also contributes to the resilience of water quality management systems in urban settings through detailed discussions on the architecture and functionality of the developed RNN model. This research addresses the development of an RNN model that can be effectively integrated with IoT systems to improve the efficiency and reliability of water quality monitoring system

The format of this paper is as follows. The review of the literature on water quality monitoring is provided in Section II. Section III provides examples of the study's methods. The experimental findings and system evaluation are covered in Section IV. The conclusion and upcoming work are finally summed up in Section V.

Literature Review

Mahesh *et al.,* (2024) investigated the level of water quality with IoT sensors and a probabilistic machine learning model. So far, Water quality in certain surroundings is analyzed using sensors depending on a number of characteristics, such as TDS, hardness, and PH level. Water quality in certain surroundings is analyzed using sensors depending on a number of characteristics, such as TDS, hardness, and PH level. The statistical machine learning model that combines the random forest model and the light gradient boost model (GBM) was used to evaluate the available data. The proposed probabilistic model has achieved an accuracy of 96.8%, a sensitivity of 94.55%, and a specificity of 98.29%. Therefore, time series-based feature extraction and the application of deep learning models for water quality assessment levels in global level data should be the main areas of future research.

Poornima *et al.,* (2024) examine a critical study of IoT and machine learning methods for water quality analysis. This study examined the various artificial intelligence (AI) methods for evaluating water quality, such as K-nearest neighbors, Deep Neural Networks, Support Vector Machines, and traditional machine learning techniques. Water resource management and spatial data analysis are done with the use of the geographical information system (GIS). Overall, the IoT system was 95 % accurate in measuring pH, Turbidity, TDS, and Temperature, while the traditional method was only 85 % accurate. For the purpose of monitoring and analyzing water quality, future work should apply deep learning technique to enhance the model accuracy.

Valadkhan, Moghaddasi, and Mohammad (2022) offered a method for predicting the groundwater quality index using LSTM-RNN and new effective parameters. The groundwater quality index's effective factors are temperature, humidity, rainfall rate, and groundwater abstraction. Monthly time series data were selected from five different locations in the Iranian area of Damavand. F-score and accuracy testing are used to assess neural network architecture in order to determine the best neural network performance. Overall, the method used was 85 % accurate. A comparison between the actual value and the prediction's result shows that, for the most part, the water quality index prediction was made correctly and logically. Future research should focus on more effective parameter optimization, including air pressure, surface evapotranspiration, and groundwater aquifer characteristics.

Ismail, Nnamdi and Ogbolumani (2023) investigated a water quality monitoring system that combines embedded IoT and machine learning. This study created a system prototype and used reliability and classifier matrices to evaluate its performance. This study considers water's physical and chemical parameters to evaluate the level of water pollutants present in drinking water. The parameters measured include temperature, pH, turbidity, Dissolved Oxygen (DO), Total Dissolved Solids (TDS), Oxidation Reduction Potential (ORP), and electrical conductivity. After analyzing the sensor data, we used Artificial Neural Network (ANN) and Support Vector Machine (SVM) machine learning algorithms to forecast the impurity level of the water measured. The performance showed that the ANN models used have the highest accuracy of 96% and are the most suitable to predict water source and status.

Pietro, Vitanio, Pietro, Domenico and Pietro (2022) focuses on an Internet of Things water quality prediction system, that remotely communicates gathered measurements leveraging the Sigfox communication technology. The proposed work demonstrates how it is possible to detect and predict water quality parameters such as pH, conductivity, oxygen, and temperature by using WaterS. The Tiziano Project dataset, and a Deep Learning algorithm based on a Long Short-Term Memory recurrent neural network were used to predict the quality of water. The system achieves accuracy of 92% and a low Mean Absolute Error of 0.20, a Mean Square Error of 0.092 , and finally a Cosine Proximity of 0.94. The obtained results are analyzed in terms of protocol suitability of the current architecture toward large-scale deployments. In general, all approaches require further development to efficiently predict high quality of water**,**

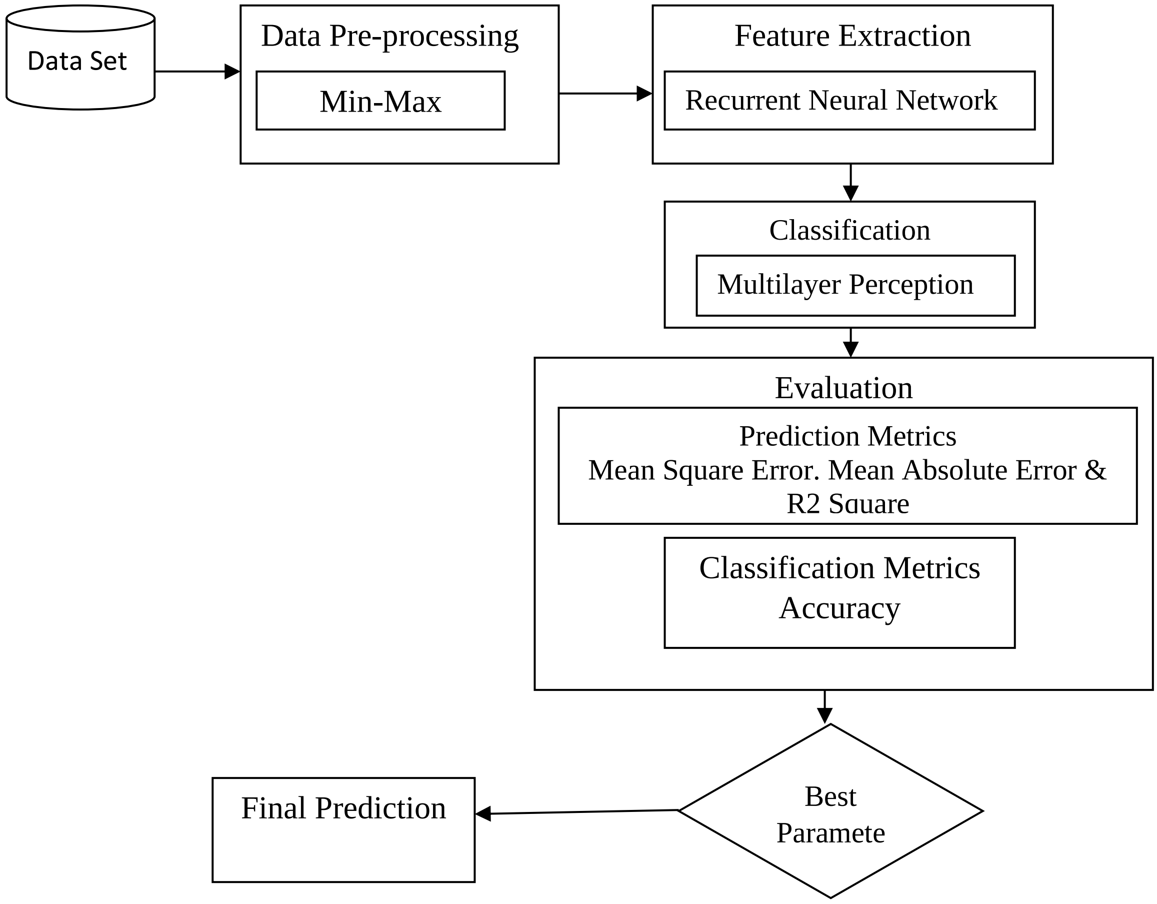
Faudzi1, Raslan, & Alias (2022) examined "IoT based real-time monitoring system of rainfall and water level for flood prediction using LSTM Network"**.** They presented an LSTM algorithm-based flood forecasting monitoring system including a rain gauge and two water level sensors is constructed. The outcomes of past data show that every model can be used to make predictions. The accuracy of the results yielded 95%. More data can be discovered for the system in subsequent work so that it can forecast and incorporate rainfall data for other time periods.

Do-Guen Yoo and Chan-Wook Lee (2021) proposed development of leakage detection model and Its utilization for RNN-LSTM Water Distribution Networks" This study examined the efficacy of a data-based leak detection model by applying it to real-world situations for a leak accident in a water distribution network system. As the leak detection model identified the majority of leak incidents more quickly than a few measuring instruments, the application result demonstrated good performance. Confusion matrix was used to assess leak detection performance, and the results consistently demonstrated greater than 90% accuracy. With smart water infrastructure being adopted, the developed model is anticipated to be a crucial software technology to proactively identify various concerns at present. To further verify the suggested leak recognition model in this work, comparative comparisons with other deep learning models should be carried out.

Methodology

Research Design

Water contamination is one of the biggest environmental problems facing humanity and the harm it causes is largely caused by a lack of emergency management. Consequently, the establishment of a suitable surveillance and early warning system to facilitate informed decision-making and water quality control is an important scientific and technical matter that needs to be resolved right away (Xiong *et al.,* 2020). In recent years, a number of machine learning techniques have made significant improvements. The proposed methodology for water quality prediction is depicted in Fig. 1. The proposed methodology aims to develop a machine learning model for water quality assessment based on a dataset containing four features: Proportion of Hydrogen (PH), Salinity, Temperature, and Specific Conductance. The dataset has already undergone preprocessing, which includes min-max imputation and data normalization.



Data Collection

Data is the foundation of this research. This study sourced data from Kaggle. Also from IoT sensors deployed across a simulated urban water distribution network. In this research, the parameters used to obtained dataset are: Proportion of Hydrogen (PH), Salinity, Temperature, and Specific Conductance. The "brisbane water quality" dataset was used in the research on water quality monitoring systems in order to train and evaluate Recurrent Neural Network models. Using this dataset primarily aims to assess the quality of water by analyzing parameters such as pH, temperature, salinity, and specific conductance. Figure 1 show the sample of the dataset

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
Timestamp	Record num	Average Wa	Average Wa	Chlorophyll	Chlorophyll	Temperatur	Temperatur	Dissolved O	Dissolved O	Dissolved O	Dissolved O	pH	pH [quality]	Salinity	Salinity [qua	Specific Cor	Specific Cor	Turbidity
8/4/2023 23:00	1468	4.834	73.484	1.621		20.018		7.472		101.175		8.176		35.215		53.262		2.068
8/4/2023 23:30	1469	2.544	106.424	1.959		19.986		7.455		100.884		8.175		35.209		53.254		1.994
8/4/2023 23:00	1470	1.26	156.755	1.62		20.001		7.43		100.571		8.171		35.207		53.252		2.03
8/4/2023 23:30	1471	0.76	281.754	1.761		19.983		7.419		100.398		8.171		35.211		53.257		1.973
8/4/2023 23:00	1472	3.397	244.637	1.635		19.986		7.429		100.538		8.171		35.208		53.253		1.944
8/4/2023 23:00	1473	1.596	100.271	1.935		19.834		7.43		100.293		8.158		35.255		53.315		2.124
8/4/2023 23:30	1474	6.622	141.844	2.103		19.829		7.435		100.354		8.158		35.259		53.321		1.95
8/5/2023 0:00	1475	9.138	52.005	1.443		19.822		7.459		100.667		8.159		35.271		53.337		2.033
8/5/2023 0:30	1476	2.982	93.117	1.433		19.804		7.446		100.471		8.166		35.273		53.339		1.995
8/5/2023 1:00	1477	9.851	53.062	1.499		19.77		7.454		100.515		8.168		35.283		53.353		1.973
8/5/2023 1:30	1478	3.37	318.489	1.083		19.741		7.447		100.371		8.167		35.29		53.362		2.014
8/5/2023 2:00	1479	0.283	324.385	1.001		19.712		7.426		100.038		8.17		35.292		53.365		1.845
8/5/2023 2:30	1480	4.989	166.41	1.435		19.711		7.404		99.745		8.169		35.29		53.363		1.861
8/5/2023 3:00	1481	1.965	6.226	1.176		19.703		7.198		96.958		8.17		35.293		53.366		2.397
8/5/2023 3:30	1482	0.647	266.369	1.287		19.701		7.204		97.023		8.174		35.284		53.355		2.34
8/5/2023 4:00	1483	8.005	219.721	1.421		19.652		4.89		65.794		8.166		35.272		53.338		3.375
8/5/2023 4:30	1484	4.907	282.524	1.286		19.634		6.78		91.198		8.168		35.276		53.343		2.213
8/5/2023 5:00	1485	4.118	139.344	1.888		19.59		7.145		96.02		8.169		35.271		53.336		2.074
8/5/2023 5:30	1486	2.691	331.264	1.427		19.585		6.646		89.32		8.168		35.267		53.331		2.498
8/5/2023 6:00	1487	6.831	336.657	1.953		19.555		7.301		98.061		8.161		35.271		53.336		1.689
8/5/2023 6:30	1488	4.434	312.831	2.47		19.539		7.022		94.291		8.164		35.278		53.346		3.742
8/5/2023 7:00	1489	5.132	82.729	2.042		19.371		6.942		92.879		8.166		35.203		53.245		2.154
8/5/2023 7:30	1490	0.611	180.177	1.865		19.367		7.276		97.333		8.162		35.166		53.195		1.628
8/5/2023 8:00	1491	7.217	352.949	1.458		19.444		7.261		97.283		8.162		35.205		53.247		1.899
8/5/2023 8:30	1492	7.251	66.444	1.082		19.451		7.221		96.735		8.163		35.154		53.178		1.756
8/5/2023 9:00	1493	4.285	90.654	1.083		19.534		7.222		96.918		8.159		35.204		53.247		1.935
8/5/2023 9:30	1494	7.336	50.886	0.934		19.611		7.257		97.568		8.155		35.271		53.336		2.145
8/5/2023 10:00	1495	6.358	51.77	1.022		19.623		7.287		97.997		8.153		35.269		53.334		2.298

Figure 1: Sample of the Dataset

Data preprocessing

Data preprocessing stage involves cleaning, formatting, normalizing, and transforming raw data into a clean and organized format suitable for analysis or modeling. It includes steps like removing duplicates handling missing values, normalization, standardization, and feature scaling. Essentially, it sets the stage for effective data analysis and modeling by ensuring data quality and compatibility. This process typically includes several tasks as stated below

Exploring the dataset

The figure 2 below shows the information about the dataset, the information such as total number of columns, total number of rows and the datatype in each row, the dataset contain 30893 rows and 20 columns.

```
data.info() # checking the general information about the dataset
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 30894 entries, 0 to 30893
Data columns (total 20 columns):
 #   Column                                  Non-Null Count  Dtype
---  ------                                  --------------  -----
 0   Timestamp                               30894 non-null  object
 1   Record number                           30894 non-null  int64
 2   Average Water Speed                     30874 non-null  float64
 3   Average Water Direction                 30893 non-null  float64
 4   Chlorophyll                             30309 non-null  float64
 5   Chlorophyll [quality]                   30086 non-null  float64
 6   Temperature                             25730 non-null  float64
 7   Temperature [quality]                   25550 non-null  float64
 8   Dissolved Oxygen                        26594 non-null  float64
 9   Dissolved Oxygen [quality]              26370 non-null  float64
 10  Dissolved Oxygen (%Saturation)          25145 non-null  float64
 11  Dissolved Oxygen (%Saturation) [quality] 24944 non-null  float64
 12  pH                                      29810 non-null  float64
 13  pH [quality]                            29586 non-null  float64
 14  Salinity                                26936 non-null  float64
 15  Salinity [quality]                      26712 non-null  float64
 16  Specific Conductance                    29527 non-null  float64
 17  Specific Conductance [quality]          29303 non-null  float64
 18  Turbidity                               28894 non-null  float64
 19  Turbidity [quality]                     28670 non-null  float64
dtypes: float64(18), int64(1), object(1)
memory usage: 4.7+ MB
```

Figure 2: Water Quality dataset information

Feature Extraction for Predicting Water Quality Monitoring System

Feature extraction is an essential step in Water Quality Monitoring System (WQMS) as it involves selecting and transforming relevant features from the prepossessed data to use in the detection of potential impurity in water. Recurrent Neural Network has embedded feature extraction function; therefore, RNN was used to extract important features. According to Moustafa and Slay (2015), during the feature extraction stage, WQMS extracts a set of discriminative features from the preprocessed data using various techniques such as statistical analysis, data mining.

Recurrent Neural Network (RNN) in Water Quality Monitoring System

RNN is a type of neural network that feeds its current state's input from its prior output. All inputs and outputs in an RNN are independent, but if the system needs to make a prediction, it needs the prior output, which means it needs to remember the previous output. Because RNNs have "memory," they can recall some details about the calculations that have been made thus far. This method uses a random neural network (RNN) to determine the final output. Once the model is trained it can be used to classify the water that is quality from the one that is not. Water Quality Monitoring feature extraction computation is shown below

The feature extraction process in RNNs is represented by the computation of the hidden state: The following function is assigned to the neurons in order to compute the prediction. Let 1….m be how many neurons there are in the system that take x1......xn +1 biased inputs and outputs:

$$\mathrm{Aw.b} = \mathrm{f}(\sum WiXi + b) \ldots\ldots\ldots\ldots\ldots\ldots \quad (3.1)$$

where the activation function is denoted by A.

$$\mathrm{A(z)} = 1/(1 + \exp(-x) \ldots\ldots\ldots\ldots\ldots\ldots \quad (3.2)$$

Classification Techniques

Classification techniques are a fundamental part of Deep learning and data mining. They are used to categorize data into different classes or categories based on input features. Multilayer perception were employed to classify the water to either Pure or impure for the specified purpose.

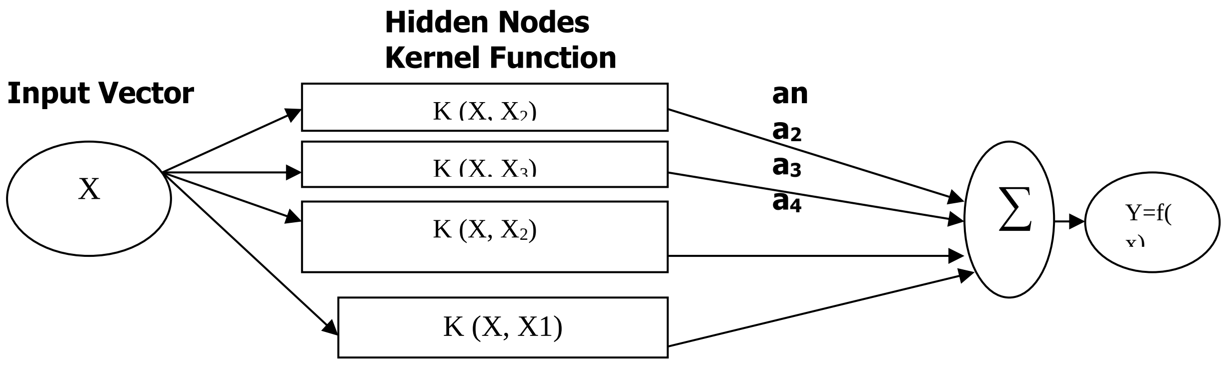


Figure 3: Architecture of the MLP neural network.

The system work follows:

Start

Step 1: Sensors measure Temperature, PH, Salinity, and Specific Conductance of the water.

Step 2: Data that has been sensed is sent to the cloud server.

Step 3: The sensed data is sent to the cloud server.

Step 4: The cloud server processes or computes the sensed data after receiving it.

Step 5: Using the training data set and the Multilayer Perception algorithm, the processed data is classified.
Step 6: Using a recurrent neural network, the classified data is transformed into the tested data set and used to forecast the range of water quality.
Step 7: The outcomes inform the authorities or individuals about the water
Stop

Performance Evaluation

The experiments are carried out using the jupyter notebook version (6.4.6). Jupyter notebook makes it easier to run and write Python scripts. It is widely used as an open-source model implementation and execution tool for AI and ML. The proposed models' performance is compared to that of numerous existing models. The classification models' performance was assessed using assessment criteria such as accuracy, Mean Square Erroe (MSE), Mean Absolute Error (MAE) and R2 Square

Accuracy

It measures the proportion of total correct predictions (both true positives and true negatives) out of all predictions made. It can be calculated from equation (3.1).

$$\text{Accuracy} = \frac{(\mathrm{TP+TN})}{(\mathrm{TP+TN+FP+FN})} \quad \ldots\ldots(3.3)$$

Mean Square Error

$$MAE = \frac{1}{\mathrm{N}} \sum_{\mathrm{i=1}}^{\mathrm{N}} \left| \mathrm{Yreal} - \mathrm{Ypred} \right| \quad \ldots\ldots(3.4)$$

Mean Square Error

$$MSE = \frac{1}{\mathrm{N}} \sum_{\mathrm{i=1}}^{\mathrm{N}} (\mathrm{Yreal\ i} - \mathrm{Y\ pred\ i})2 \quad \ldots\ldots(3.5)$$

R^2

$$\mathrm{R2} = 1 - \frac{\sum \mathrm{i} = 1(yreal - ypred)^{2}}{\sum i = 1\ (yreal - y)2} \quad \ldots\ldots(3.6)$$

Result and Discussion

Performance of RNN model with temperature as indicator

Table 1: Performance evaluation of RNN using temperature as indicator

Model	Mean Square Error (MSE)	Mean Absolute Error (MAE)	R2 Square
LSTM	1.33	0.81	0.85

The average squared difference between the actual and anticipated values is measured by MSE. An MSE of 1.33 in this instance means that the squared difference between the actual and anticipated temperatures is, on average, 1.33 units. Better model performance is indicated by lower MSE values. The model appears to be reasonably reliable in predicting temperature. The model's predictions are often within a reasonable range of the actual temperature values, as indicated by the MAE of 0.81. An R2 value of 0.85 means that the model can account for 85% of the variability in the target variable. Given its high value, the RNN model appears to have a high degree of predictive power and the ability to identify a sizable percentage of the underlying patterns in the data. Therefore, it can be concluded that the model is doing a good job of forecasting temperature in the context of water quality monitoring due to its low MSE (1.33), low MAE (0.81), and high R2 (0.85).The model accounts for a significant amount of the temperature variability, and the predictions are typically accurate and near the genuine values. Figure 7 shows the forecasting visualization for the temperature in the dataset

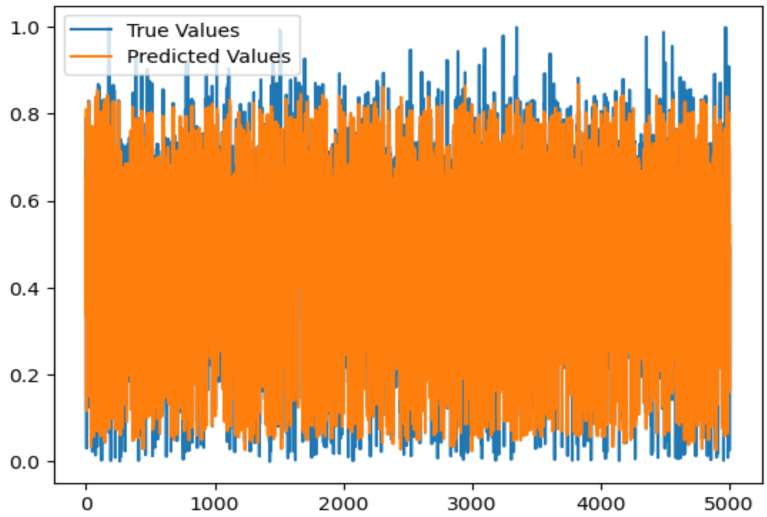


Figure 7: Forecasting Visualization for the Temperature in the Dataset

The figure 7 visualizes the true values and the forecast values of the temperature indicator on water quality dataset. The blue line is the true values while the orange line is the predicted values on training set (80%)

Performance of RNN model with pH as indicator

Table 2: Performance evaluation of RNN using pH as indicator

Model	**Mean Square Error (MSE)**	**Mean Absolute Error (MAE)**	**R2 Square**
LSTM	0.00044	0.012	0.95

The model is very good at predicting pH levels in water quality monitoring due to its combination of an exceptionally low MAE (0.012), extremely low MSE (0.00044), and high R^2 (0.95). With relatively little error and a great explanatory power for the pH variability, the forecasts are remarkably precise. This suggests that the water quality system's pH monitoring can be done with the model in a reliable way.

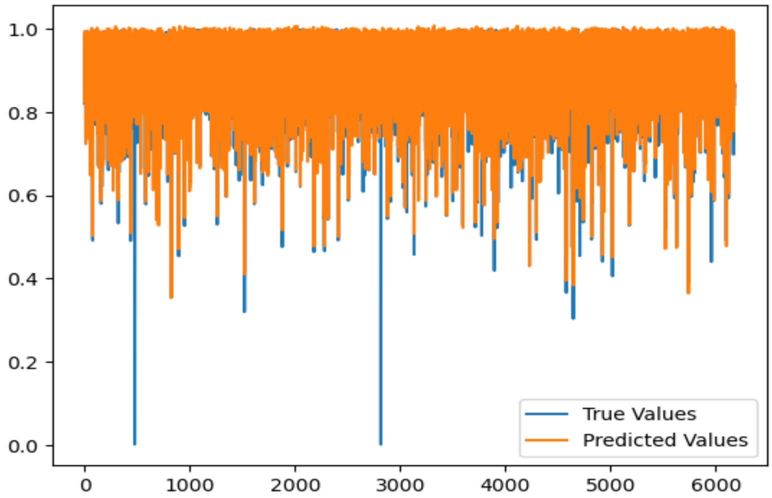


Figure 8: Forecasting Visualization for the pH in the Dataset

Figure 8 visualizing the true values and the forecast values of the pH indicator on water quality data set with 50 epochs. The blue line is the true values while the orange line is the predicted values on training set (80%)

Performance of RNN model with Salinity as indicator

Table 3: Performance evaluation of LSTM using Salinity as indicator

Model	Mean Square Error (MSE)	Mean Absolute Error (MAE)	R2 Square
LSTM	0.25	0.22	0.97

The model's predictions are, on average, only 0.22 units off from the actual salinity levels, with an MAE of 0.22. With relatively minor variations from the true values, the model's salinity predictions are consistently accurate, as evidenced by the low mean absolute error (MAE). With an R2 value of 0.97, the model accounts for 97% of the variation in salinity. With a high R2 value, the model appears to have a good fit to the data, which indicates that it accurately depicts the correlation between salinity and the input characteristics. Due to the low MAE (0.22), low MSE (0.25), and very high R2 (0.97), it can be concluded that the model is doing a remarkable job of forecasting salinity in water quality monitoring. The model explains almost all of the salinity variations, and the forecasts are very accurate with few errors. This robust performance suggests that the model for monitoring salinity in a water quality system is dependable and efficient.

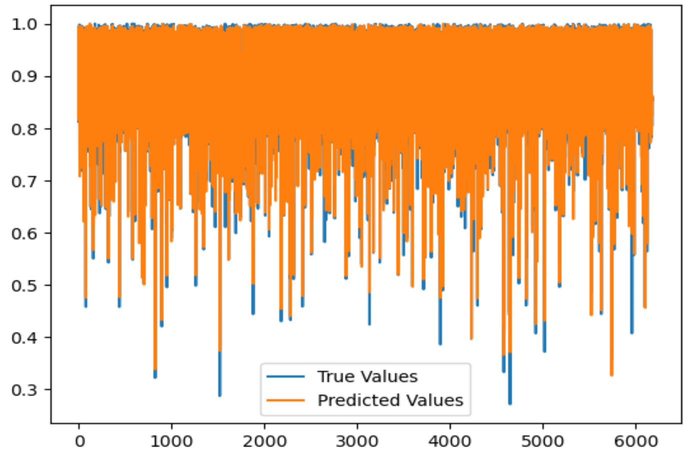


Figure 9: Forecasting Visualization for the Salinity in the Dataset

The figure 9 visualizes the true values and the forecast values of the Salinity indicator on water quality data set with 50 epochs. The blue line is the true values while the orange line is the predicted values on training set (80%)

Performance of RNN model with Specific Conductance as indicator

Table 4: Performance evaluation of RNN with Specific Conductance

Model	Mean Square Error (MSE)	Mean Absolute Error (MAE)	R2 Square
LSTM	1.72	0.43	0.93

The model's predictions are, on average of MSE 1.72 and MAE of 0.43 units off from the actual conductivity values. Although this shows strong predictive performance, when compared to other metrics, the errors are marginally more substantial. R^2 measures the proportion of variance in the conductivity data explained by the model. An R^2 value of 0.93 means that 93% of the variance in conductivity is captured by the model. This high R^2 suggests that the model is highly effective at explaining the relationship between the input features and conductivity, although there is still 7% of the variance unexplained.

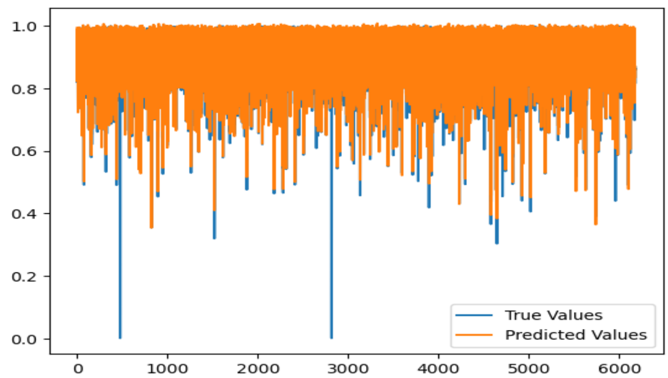


Figure 10: Forecasting Visualization for the Specific Conductance in the Dataset

The Specific Conductance indicator's actual and predicted values for a 50-epoch water quality data set are shown in figure 10. The orange line represents the training set's anticipated values (80%), whereas the blue line represents the actual values.

Analysis of RNN model results with all the indicators used

Table 5 compares the performance metrics four 4 parameter: Temperature, pH, Salinity, and Specific Conductance.

Table 5: Analysis of performance metrics on RNN model

Indicators	MSE	MAE	R2
Temperature	1.33	0.81	0.85
pH	0.00044	0.012	0.95
Salinity	0.25	0.22	0.97
Specific Conductance	1.71	0.43	0.93

The performance metrics of the RNN model with all of the indicators utilized for the water quality monitoring system are displayed in Table 5. The model performs well with low Mean Square Error (MSE), low Mean Absolute Error (MAE), and high R2 for all the parameters in the above table that are thought to be indicators of water quality. This suggests that the model is accurate in predicting, particularly pH values, minimizing both large errors and overall prediction errors. The model's prediction accuracy for pH is better, as indicated by lower MSE and MAE while the model is better at explaining the variation in salinity values, as indicated by the high R^2

Comparison with Previous Work in the Literature

In this section, the results of the research are compared with the previous studies that were reviewed in the literatures. Table 4.2 shows that the results of the proposed model are better than the results obtained from the previous studies.

Table 4.2: The results of the proposed model

S/N	Author(s)	Title	Methodology	Result
1	Mahesh et al., (2024)	investigated the level of water quality with IoT sensors and a probabilistic machine learning model	The statistical machine learning model that combines the random forest model and the light gradient boost model (GBM) was used to evaluate the available data.	The proposed probabilistic model has achieved an accuracy of 96.8%, a sensitivity of 94.55%, and a specificity of 98.29%.
2	Valadkhan, Moghaddasi, and Mohammad (2022)	offered a method for predicting the groundwater quality index using LSTM RNN and new effective parameters	Designing and optimizing a Convolutional Neural Network (CNN) with a Pelican Optimization Algorithm (POA) and RNN based on water quality prediction	Overall, the method used was 85 % accurate.
3	Proposed work	water quality control system	The following machine learning models were trained and evaluated	The proposed probabilistic model

	using recurrent neural network	on the brisbane water quality datasets by analyzing parameters such as pH, temperature, salinity, and specific conductance	has achieved an accuracy of 98.1%,

Conclusion

An artificial neural network class called RNNs is made specifically to manage sequential data. Recurrent neural networks (RNNs) are especially well-suited for time-series data analysis, such as that produced in water distribution systems (WDS), since they can handle input sequences using their internal state, or memory, in contrast to classic neural networks. The effectiveness of Temperature, pH, Salinity and specific conductance on RNN model were investigated in this paper. The best indicator for water quality monitoring system was obtained by applying these approaches to the water quality dataset.

In this study, we suggested an Internet of Things (IoT) system for fisheries and aquaculture water quality monitoring, with a focus on the model for predicting quality indicators including salinity, pH, temperature, and specific conductance. Results from experiments conducted on the four indicator data sets under consideration demonstrate that the suggested method is applicable to the actual system. The technology can assist farmers in managing water quality and hence harvesting both quality and quantity of shrimp and fish by tracking these real-time indicators and providing early warning. The RNN model is recommended for use in an IoT environment's water quality prediction system because to its superior efficiency and accuracy, particularly when it comes to the pH indicator, as compared to certain other models.

References

Ismail A. A., Nnamdi I. N., & Omolola A. O. (2023). A hybrid machine learning and embedded IoT-based water quality monitoring system. *Environmental Science and Pollution Research*, 1-25, https://doi.org/10.1016/j.iot.2023.100774

Liu, P., Wang, J., Sangaiah, A. K., Xie, Y., & Yin, X. (2019). Analysis and prediction of water quality using LSTM deep neural networks in IoT environment. *Academia journal of sustainability*, 11(7), 2058.

Mahesh, T. R., Surbhi, B. K., Balajeea, A. A., Thippa, R. G., Eid, A., & Vinoth, K. V. (2024) Water quality level estimation using IoT sensors and probabilistic machine learning model. *Journal of British Hydrological Societies. Hydrology Research 55*(7). doi: 10.2166/nh.2024.048

Moonsun, R. K., Barzegar, R., Aalami, M. T. & Adamowski, J. (2022). Short-term water quality variable prediction using a hybrid CNN–LSTM deep learning model. *Stochastic Environmental Research and Risk Assessment*, *34*(2), 415-433.

Nikhil, T. N., Kumar, N. & Koditala N. (2018). Deep learning approach for forecasting water quality in IoT systems. *International Journal of Advanced Computer Science and Applications*, *11*(8), 686-693.

Pietro, B., Vitanio, D., Pietro D. G., Domenico L., & Pietro, T. (2022). Water quality prediction on a Sigfox-compliant IoT device: The road ahead of WaterS. Ad Hoc Networks Volume 126, 1, 102749. https://doi.org/10.1016/j.adhoc.2021.102749

Poornima, T., Kothalam, M., Nagarajan, R., Pachaivannan, T., & Vallidevi, L. (2024). Critical review on water quality analysis using IoT and machine learning models. *International Journal of Information Management Data Insights*. www.elsevier.com/locate/jjimei

Sana, A. F., Narwane, V. S., & Gardas, B. B. (2022). Contribution of iternet of things in water supply chain management: A bibliometric and content analysis. *Journal of Modelling in Management*, *18*(2), 549-577.

Smith, C. W., & Brown, D. G. (2019). Development of leakage detection model and its application for water distribution networks using RNN-LSTM. *Academia journal of Sustainability*, *13*(16), 9262.

Smith, S., & Brown, P. (2019) IoT enabled monitoring and control of water distribution network:(084). In *WDSA/CCWI Joint Conference Proceedings* (Vol. 1).

Theofanis, N., Christos, N. & Marios, N. (2014). Deep learning approach for forecasting water quality in IoT systems. *International Journal of Advanced Computer Science and Applications*, *11*(8), 686-693.

Tsotsope, B., Wu, D., Tsu, K., Chi, H., & Zhang, C. (2020). Framework for the implementation of an Internet of Things (IoT)-based water distribution and management system. *Clean Technologies and Environmental Policy*, *23*, 271-283.