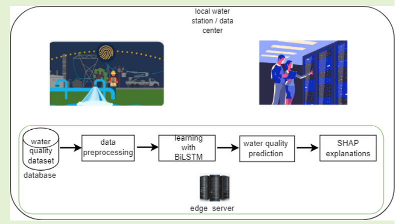
Water Quality Assessment Tool for On-site Water Quality Monitoring

Segun O. Olatinwo, Trudi H. Joubert, *Senior Member, IEEE*, and Damilola D. Olatinwo

***Abstract*—Reliable water quality monitoring requires on-site processing and assessment of water quality data in near real-time. This helps to promptly detect changes in water quality, prevent biodiversity loss, safeguard the health and well-being of communities, and mitigate agricultural problems. To this end, we proposed a Highway-Bidirectional Long Short-term Memory (Highway-BiLSTM)-based water quality classification tool for potential integration into an edge-enabled water quality monitoring system to facilitate on-site water quality classification.**

The performance of the proposed classifier was validated by comparing it with several baseline water quality classifiers. The proposed classifier outperformed the baseline water classifier in terms of accuracy, precision, sensitivity, F1-score, and confusion matrix. Specifically, the proposed water classifier surpassed the random forest (RF) classifier with 2% accuracy, precision, sensitivity, and F1-score. Moreover, the proposed classifier achieved an increase of 4% in accuracy, precision, sensitivity, and F1-score for classifying water quality compared with the Gradient Boosting classifier. Additionally, the proposed method has 4% increase in accuracy, sensitivity, F1-score, and 3% increase in precision compared to the support vector machine (SVM) water quality classifier. The proposed method outperformed the artificial neural network (ANN) classifier by 1% accuracy, precision, sensitivity, and F1-score. Finally, the proposed method demonstrated rare errors in accurately classifying complex water quality samples. These findings suggest that our proposed method could be used to effectively classify water quality to aid accurate decision making and environmental management.



***Index Terms*—environmental monitoring, water quality monitoring, marine biodiversity preservation, water pollution control, public health protection, agricultural productivity**

I. Introduction

WATER is essential for the survival, growth, and well-being of humans and other living things. Humans require clean water to remain hydrated and produce goods and services. Plants and animals also depend on water for survival, growth, and reproduction [1]. Water promotes the health and well-being of all living things. These examples illustrate the crucial role of water in life. However, water quality has declined over the years due to defective water infrastructure, anthropogenic activities (e.g., industrialization and urbanization), and neglect [2], [3]. These factors have caused water quality to fall below the standards recommended by international bodies such as the World Health Organization (WHO) [4]. For instance, according to a report by the South African government [5], the water quality across South African provinces has dropped significantly owing to the poor condition of the water supply systems.

Furthermore, due to declining water quality, about two billion people (26% of the world population) lack access to safe water [6], and approximately three billion people suffer from water-related diseases, such as cholera, typhoid, and dysentery every year [7]. These statistics indicate the need for reliable water quality monitoring methods. From a technical point of view, the issues identified affect the physical, chemical, and biological properties of water, thus making it unclean for several purposes related to drinking, production, irrigation, marine ecosystems, etc. [8].

Key organizations such as the WHO, United Nations, and United States Environmental Protection Agency (USEPA) have identified in their strategic report plans that monitoring water infrastructure including sources and supply systems is a priority. Classifying water quality, detecting changes in water quality, controlling pollution (e.g., green tides) in water environment, preventing the supply of polluted water, and improving the quality of life of humans and marine mammals are also key priorities. These priorities have attracted researchers from academia and industry to investigate the development of solutions that can be used to monitor and classify water quality in near real-time [9].

Water quality monitoring is mostly conducted using a traditional approach, wherein environmental agencies collect physical parameters (such as pH and temperature), microbiological indicators (including faecal and total coliforms), and chemical measurements (such as dissolved oxygen and nitrate levels) from water samples. These samples are then transported to

Manuscript received October, 2023.

S.O. Olatinwo is with the Department of Electrical, Electronic and Computer Engineering, University of Pretoria 0002, Pretoria, South Africa (Corresponding author: segunlatinwo@gmail.com).

T.H. Joubert is with the Department of Electrical, Electronic and Computer Engineering, University of Pretoria 0002, Pretoria, South Africa. (e-mail: trudi.joubert@up.ac.za).

D.D. Olatinwo is with the Department of Electrical, Electronic and Computer Engineering, University of Pretoria 0002, Pretoria, South Africa (Corresponding author: damibaola@gmail.com).

This article has been accepted for publication in IEEE Sensors Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2024.3383887

distant laboratories for analysis, aiming to assess the suitability of water sources. However, the conventional method of water quality monitoring is both costly and ineffective, as many water quality parameters are better analyzed in-situ.

To address the limitations of traditional water quality monitoring methods, researchers have explored various data-driven approaches for classifying water suitability. However, these methods currently face challenges related to data preprocessing, lack the ability to handle complex datasets, and suffer from limited interpretability in water quality classification. Additionally, they do not support on-site processing and classification of water quality data due to the constrained computational resources of water quality devices within water quality monitoring systems. As a result, on-site computational tasks are infeasible [10]. Consequently, existing water quality devices merely collect data from water bodies and transmit it to remote monitoring centers for analytics and processing. Unfortunately, this approach introduces delays in data analysis, making it challenging to classify water quality and promptly detect changes in near real-time.

Given the current limitations associated with data-driven methods for water quality classification, achieving accurate and reliable results remains an open problem [11], [12]. Consequently, ongoing water research primarily focuses on exploring novel water quality classifiers to overcome the limitations of existing approaches [11], [12].

To address the research gaps in the literature, this study focused on the development of an efficient water quality classification tool for processing and classifying water quality in situ. The main contributions of this study are as follows.

- We introduced the integration of edge computing with water quality monitoring systems to facilitate on-site water quality data analytics, processing, and classification.
- We addressed the data preprocessing problems faced by the existing water quality classifiers using a robust scaler technique.
- We addressed the class imbalance and data scarcity problems associated with water quality data using SMOTE data augmentation.
- We introduced the use of the bidirectional long short-term memory (BiLSTM) to enable the processing and classification of complex water quality data.
- We introduced the use of a highway network mechanism to improve the computational time complexity problem of deep learning-based water quality classifiers.
- We used a computational efficient random search optimization method to select the best model for the proposed water quality classifier.
- We provided an interpretability mechanism to enable the explanation of the classification of water quality.

We organized the content of this article follows. Section II presents a review of the related studies. Section III presents the methods used in this study. In Section IV, we present results and discussion on the performance evaluation results of the proposed method. Section V concludes this study.

II. Related Studies

In the literature, several methods have been employed to classify water quality samples and determine their suitability for various purposes [13]–[17]. Ladjal et al. [13] utilized artificial neural networks (ANN) and support vector machines (SVMs) to assess the quality of water from the Tilesdit dam. They considered parameters such as electrical conductivity, temperature, pH, and turbidity, categorizing water quality as either mediocre, moderate, or excellent. The ANN water quality classifier achieved a training accuracy of 99.75% and a testing accuracy of 99.13%. Similarly, the SVM classifier demonstrated a training accuracy of 99.00% and a testing accuracy of 98.48%. However, the min-max scaler data transformation method used during data preprocessing is insufficient for handling skewness in water quality data, especially in the presence of outliers. Additionally, the issue of data scarcity, common in water quality datasets, was not addressed. While machine learning algorithms like ANN and SVM offer simplicity, they struggle with non-linear relationships, posing challenges in capturing complex structures. Furthermore, the work of Ladjal et al. [13] did not provide an interpretability mechanism for their model classifications.

Khelil et al. [14] compared the performance of SVM and the Long Short-Term Memory (LSTM) binary classifiers using the Tilesdit dam water quality data. They reported an accuracy of 99.81% for the SVM classifier and 98.70% for the LSTM water quality classifier. However, these classifiers have limitations. The SVM classifier lacks the capability to process complex water quality data and capture long-distance dependencies due to the curse of dimensionality. Consequently, it may struggle to perform well in practical scenarios when confronted with intricate water quality data. On the other hand, LSTM, while effective, is a data-intensive method that requires a large dataset to perform optimally. Additionally, it faces challenges related to computational time complexity. Unfortunately, the issue of water quality data scarcity was not addressed in their study. Furthermore, Khelil et al. [14] did not provide an interpretability mechanism for their model classifications.

Dilmi and Ladjal [15] proposed multiclass water quality classifiers to assess the water quality status of the Tilesdit water quality data, categorizing it into three classes: class one, class two, and class three. They employed feature extraction methods, such as Principal Component Analysis (PCA) , Linear Discriminant Analysis (LDA), and Independent Component Analysis (ICA), to develop SVM and LSTM classifiers. However, despite their high performance, these classifiers lacked an interpretability mechanism, making it challenging to explain how their decisions were made. Furthermore, they did not consider the computational constraints (training time and testing time) associated with the LSTM algorithm. Additionally, the issue of water quality data scarcity remained unaddressed. Lastly, the LSTM method's effectiveness is constrained by the requirement for a large dataset.

Abuzir and Abuzir [16] used PCA to obtain low-dimensional water quality data. Then, they used simple naive Bayes, multilayer perceptron (MLP), and J48 methods to classify

This article has been accepted for publication in IEEE Sensors Journal. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2024.3383887

water quality data based on organic carbon, pH, hardness, solids, chloramines, turbidity, conductivity, trihalomethanes, and sulfate levels. They reported a precision of 0.63, recall of 0.87, and F1-score of 0.74 for the naive Bayes model. In addition, they reported a precision of 0.63, recall of 0.92, and F1-score of 0.75 for the J48 model. Moreover, they reported a precision of 0.67, recall of 0.86, and F1-score of 0.75 for the MLP model. However, naive Bayes, MLP, and J48 are simple methods, which lack the ability to effectively process nonlinear and complex water quality data. Furthermore, the naive Bayes, multilayer perceptron (MLP), and J48 water quality classifiers lacked the capability to account for more important water quality features. In addition, their model water quality classification lacked important explanations to understand how the model arrived at a decision.

Khan et al. [17] employed the PCA feature extraction method to reduce the dimensionality of water quality features, including electrical conductivity, suspended solids, chemical oxygen demand, chloride, alkalinity, pH, turbidity, and dissolved oxygen. The resulting low-dimensional data were then utilized to develop multiclass water quality classifiers using three different algorithms: SVM, Random Forest, and AdaBoost. These classifiers were applied to classify the water quality status of the Gulshan Lake data into five distinct classes: unsuitable for drinking, bad, medium, good, and excellent. However, these classifiers have various limitations. The min-max scaler method used for data transformation lacked the ability to effectively rescale water quality data in the presence of outliers, which is a common issue in water quality datasets. The problem of data scarcity associated with water quality data was not addressed. The authors did not provide any interpretability mechanism to explain the decisions made by their water quality classifiers. The two-way holdout method used during training and testing may suffer from data leakage, potentially introducing variance and bias to the classifiers' performance. This could impact their generalization ability when deployed on unseen data.

The results of the existing water quality classifiers demonstrate progress in leveraging computational methods to enhance water quality monitoring and pollution control. However, most of these classifiers encounter challenges related to data preprocessing. For instance, the min-max scaler method, employed in studies by Ladjal et al. [13] and Khan et al. [17], lacks robustness when dealing with outliers in water quality data. Consequently, it is inefficient in reducing skewness in the transformed data. Such skewed data can potentially impact the quality of training data and the performance of trained water quality classifiers.

Another critical issue is data scarcity in systemic water quality monitoring using water quality data. Unfortunately, many existing studies on water quality classification (e.g., [13]–[16], [18]) have overlooked this problem. Addressing data scarcity is essential for the development and implementation of reliable water quality classifiers.

Many existing classifiers rely on machine learning methods, as documented in references [13]–[16], [18]. These methods employ simple equations and often assume a linear relationship between input and output variables. However, water quality data exhibit complex and nonlinear relationships between the water quality parameters that simple machine learning methods may not fully capture or account for. Furthermore, water quality data exhibit long-term dependencies. Moreover, traditional machine learning approaches are ill-suited for handling high-dimensional data due to the curse of dimensionality. To address this limitation, some water quality classifiers have turned to unsupervised machine learning techniques like PCA and ICA. These methods aim to transform high-dimensional water quality data into lower-dimensional representations for dimensionality reduction. Unfortunately, reduced data may occasionally omit crucial features. Consequently, despite the performance of existing machine learning-based water quality classifiers, they may struggle in practice when confronted with highly complex and nonlinear water quality data.

Existing water quality classifiers based on deep learning utilize Long Short-Term Memory (LSTM) networks to encode water quality feature sequences from the front to the back. However, LSTM models lack the ability to encode input water quality feature sequences in reverse, which hinders the extraction of bidirectional contextual information for deep understanding and improved model performance. As a consequence, a significant portion of crucial underlying water quality information remains unexplored. Furthermore, LSTM architectures exhibit data intensity and demand large volumes of water quality data for optimal performance. Additionally, they may suffer from computational time complexity issues when efficient information flow optimization mechanisms are absent. Consequently, there is a pressing need for a more reliable model that can be effectively deployed in near real-world scenarios.

Existing water quality classifiers are primarily suited for processing and classifying the status of offline water quality data collected from remote water bodies using Internet of Things (IoT) water quality sensors (such as pH sensors and dissolved oxygen sensors). However, these classifiers are not well-suited for near real-time, on-site processing and classification of online water quality data. Implementing on-site water quality data processing and classification systems would significantly enhance the provision of near real-time early warnings.

Existing water quality classifiers often lack an interpretability mechanism. Consequently, they operate as black boxes, making it challenging to comprehend their decision-making process. The concept of interpretability aims to provide users with a clearer understanding of how different water quality parameters impact the output of a water quality classifier. By addressing this need, there is an opportunity to develop novel water quality classifiers that offer greater interpretability.

To overcome the limitations of existing water quality classifiers, we employed a multifaceted approach. First, we utilized an advanced robust scaler method to address data transformation challenges encountered by prior classifiers. Next, we applied the Synthetic Minority Over-sampling Technique (SMOTE) to tackle class imbalance and data scarcity issues inherent in water quality data. We utilized a Bidirectional Long Short-Term Memory (BiLSTM) network to model and capture the complex and nonlinear patterns in water quality data. Additionally, we employed this method to capture long-term dependencies within the same data. Our choice

This article has been accepted for publication in IEEE Sensors Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2024.3383887

of deep learning methods aimed to address the limitations of water quality classifiers associated with machine learning approaches. Specifically, we focused on: (1) processing and capturing the complex and nonlinear patterns in the water quality data using bidirectional information, and (2) effectively capturing long-distance dependencies in the data. The bidirectional information also facilitates capturing dependencies across distant data points. Our approach aimed to improve model generalization and enhance classification performance. Furthermore, we incorporated a highway network mechanism to address the computational efficiency problem associated with deep learning-based water quality classifiers. In addition, we introduced the integration of an edge computing technique with water quality monitoring systems, enabling in-situ water quality data processing and classification in near real-time. Finally, our approach emphasized the use of an explainability mechanism, resulting in the development of an interpretable water quality classifier. These enhancements collectively contribute to a more reliable and effective model for practical applications in real-world scenarios.

III. Methods

This section provides comprehensive insights into the global schematic diagram of the proposed water quality monitoring system. It delineates various crucial aspects, including the water quality data collection process, subsequent data processing steps, the novel water quality classification method, baseline classification approaches, and the evaluation metrics employed to assess the effectiveness of these classification methods.

A. Edge Enabled Water Quality Monitoring System

The global schematic architecture of the proposed edge-enabled water quality monitoring system is presented in Fig. 1. It consists of four layers: the IoT water quality sensor layer, local base station (BS) layer, edge computing layer, and water quality classification layer.

The IoT water quality sensor layer is composed of the water bodies and the IoT water quality sensors. The water body represents the interested water environment to be monitored to control water pollution and prevent the distribution of unsafe water for various consumption purposes. The IoT water quality sensors are intended to collect important water quality data such as pH and fecal coliforms [4], [18]–[20] from the water bodies and forward the collected data through an IoT communication technology to the local base stations at the water site(s).

The local BS layer consists of the BSs which are used to coordinate and aggregate the water quality data collected by the IoT water quality sensors. The BSs are integrated with the edge computing layer for data computation and storage purposes. The edge computing layer consists of two edge nodes, each with an edge server and a database. The layer performs data analytics, computation, and storage functions in the proposed water quality monitoring system. The servers are used for the local processing of the collected water quality data, and running of an integrated Highway-BiLSTM-based water quality classifier that is proposed for water quality classification.

The water quality classification layer uses the proposed Highway-BiLSTM model to determine the quality of water quality samples based on the water quality features. The explanations of water quality classifications are provided using an interpretabilty (SHAP) mechanism that we introduced. These interpretations from the proposed method's classification can be leveraged to detect changes in water quality. This can help to facilitate prompt decision-making that can contribute to environmental protection, mitigation of health issues, prevention of biodiversity loss, and efficient management of water resource allocation. The architecture for the proposed edge-enabled water quality monitoring system is provided in Fig. 1.

The proposed edge-enabled water quality monitoring system was motivated by the lack of support for near real-time on-site processing of water quality data, online water quality data classification, and the explainability of classifications in existing literature. Existing traditional water quality monitoring systems, deployed at monitoring stations or sites, transmit collected data to remote water quality monitoring centers (where analysis, processing, and classification take place) via API cloud servers. However, this approach can hinder near real-time monitoring of water parameter changes due to the distance and delay involved in data analysis, processing, and classification. Consequently, both ecosystem health and public health may be compromised [21]–[23].

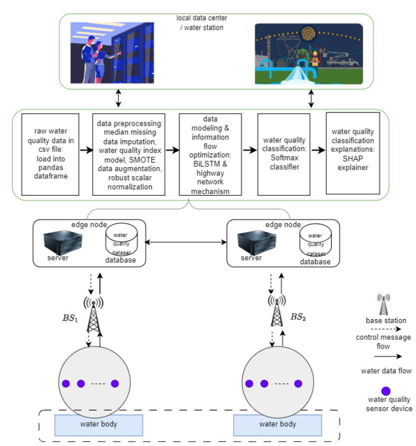


Fig. 1: Global schematic architecture of the proposed edge-enabled water quality monitoring system

B. Problem Formulation

In this study, we consider a dataset comprising of M water quality samples. Each sample is represented by a sequence of n numerical features that are dependent on each other. Our goal is to safeguard human health, plant health, and aquatic ecosystems. To achieve this, we monitor various water quality parameters, including fecal coliform, total coliform, pH, temperature (temp), conductivity, biological oxygen demand (BOD), dissolved oxygen, and nitrate (NI).

For each water quality sample, we denote the feature vector as $X_m = \{x_m^1, x_m^2, ..., x_m^n\}$, where x_m^i represents the $i-th$ feature of the $m-th$ sample. Additionally, we assign an

This article has been accepted for publication in IEEE Sensors Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2024.3383887

associated class label to each sample, denoted by y_m. The label space consists of five water quality status categories: excellent, good, poor, very poor, and unsuitable. That is, $y_m \in \{1, 2, 3, 4, 5\}$ is the class label associated with each $m-th$ sample.

We approach this as a multi-class classification problem and employ a deep learning technique. Our model outputs probabilities that a given water quality data sequence belongs to a specific water quality status. Ultimately, the predicted water quality class informs decisions regarding the appropriate utilization of these vital water resources.

C. Data Source

We utilized a publicly available water quality dataset that encompasses crucial measurements relevant to the focus of our study [24]. Specifically, we employed the Indian water quality dataset, which was collected from diverse water bodies across various Indian states during the period from 2005 to 2014. This dataset was made publicly available and stands out due to its inclusion of essential physical parameters (such as pH, electrical conductivity, and temperature), microbiological indicators (including fecal and total coliforms), and chemical attributes (such as biological oxygen demand and dissolved oxygen). These features render it well-suited for water quality monitoring and pollution control. The dataset contains 1,780 samples.

D. Data Preprocessing

To enhance the understanding of water quality data and prepare it for modeling, we conducted a series of essential data preprocessing steps. These tasks were crucial to ensure accurate analysis and effective modeling. The steps included analyzing missing data, examining data distributions, computing water quality indices, assessing the impact of input features, studying class distribution, applying data transformations, and partitioning the data for subsequent modeling purposes. By systematically addressing these aspects, we created a high quality dataset ready for further analysis and classification tasks.

1) Missing Data Analysis: Missing data analysis is an important step in data preprocessing since missing values can results into bias and under-representation (or inaccuracies) in statistical analysis and machine learning models [25]. Missing data is a common problem in environmental monitoring fields. It can occur due to sensor failures, data transfer issues from sensors, poor strategies for water quality data collection, and changes in water monitoring sites [18]. Hence, we performed missing data analysis to understand and handle missing values in the data. It was observed from Fig. 2 that fecal coliform had the most missing values, whereas pH had the least missing values. Following the missing data analysis, we addressed the missing data problem in the data to improve the statistical power of the data analysis and data modeling. We used the median method to handle the missing data problem in the dataset.

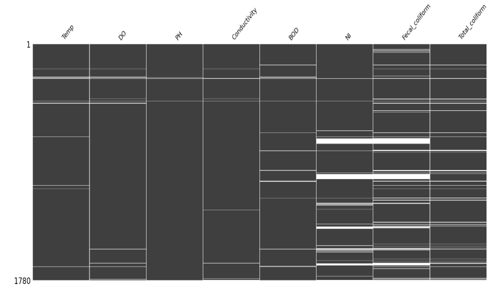


Fig. 2: Missing values in the water quality dataset

2) Data Distribution Analysis: We analyzed the water quality data features to understand their distribution. This helped to identify the features with a skewed distribution and any outliers or errors. This understanding of the dataset distribution was used to determine a suitable data transformation method to normalize the data and improve the quality of the data.

We used density plot and box plot visualization tools available through the Python seabon and matplotlib libraries to visualize the distribution of the water quality dataset in Figures 3 and 4.

The density plots indicate that all features are far from a normal distribution and are characterized by either a left-skewed distribution or a right-skewed distribution due to the presence of outliers possibly resulting from environmental factors [26], [27] and technical errors [28], [29].

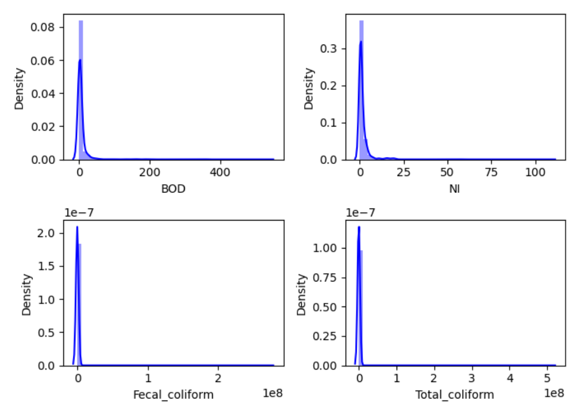


Fig. 3: Density plots of the distribution of the water quality dataset's BOD, NI, fecal coliform, and total coliform features

3) Computation of Water Quality Indices: We utilized the weighted arithmetic index method as an analytical tool to compute water quality indices (WQIs) for the water quality samples similar to Khan et al. [17]. The input parameters for the WQI calculation include temperature, DO, conductivity, BOD, NI, pH, fecal coliforms, and total coliforms. This method combines physical, microbiological, and chemical water quality parameters, assigning appropriate weights to each parameter. The resulting single numerical value (WQI) represents the overall water quality. The WQI serves as a valuable tool for assessing and comparing water quality across different samples or locations. To calculate the WQIs, we applied Eqns. 1–4:

This article has been accepted for publication in IEEE Sensors Journal. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2024.3383887

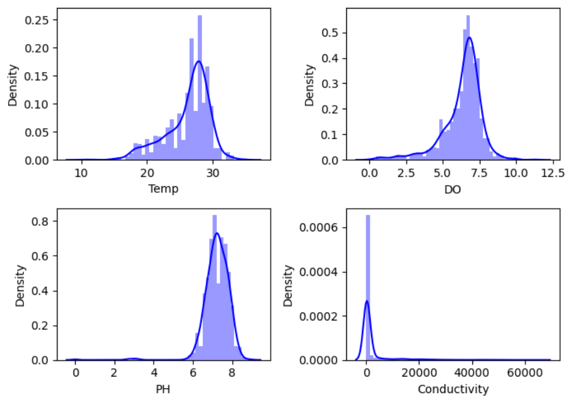


Fig. 4: Density plots of the distribution of the water quality dataset's temperature, DO, pH, and conductivity features

$$WQI = \frac{\sum_{l=1}^{L} q_l \times w_l}{\sum_{l=1}^{L} w_l} \tag{1}$$

where L represents the total number of water quality parameters considered in the calculation of the WQI, q_l describes the scale of the quality rating for each water quality parameter l determined by using Eqn. 2, and w_l represents the unit weight calculated in Eqn. 3 for each water quality parameter.

$$q_l = 100 \times \left(\frac{m_l - m_{ideal}}{r_l - m_{ideal}} \right) \tag{2}$$

where m_l represents the measured value of water quality parameter l, m_{ideal} denotes the ideal value of water quality parameter l in pure water (e.g, pH = 7.0, DO = 14.6 mg/l, and 0 for other parameters), r_l represents the standard value recommended for parameter l by the key water quality regulation bodies like the World Health Organization (WHO) [4], [18], [19]. The recommended permissible limits for water quality parameters are listed in Table I.

$$w_l = \frac{K}{r_l} \tag{3}$$

In (3), K represents the constant of proportionality. K can be calculated by applying Eqn. 4:

$$K = \frac{1}{\sum_{l=1}^{L} r_l} \tag{4}$$

The ranges of the WQI categories and their respective water quality statuses and applications are presented in Table II.
The WQI parameter in Table II indicates that there are five classes of water quality status in the data. They include "excellent", "good", "poor", "very poor", and "unsuitable". The implications of the water quality grading in Table II are described as follows:

- Excellent (0-25): Water quality in this range is clean, posing minimal risk to human health, plant life, and ecosystems. It ensures a healthy environment for aquatic species.
- Good (26-50): Water quality remains high, benefiting humans, plants, and marine animals. It supports a thriving ecosystem and sustains aquatic life.
- Poor (51-75): Water quality in this range is primarily suitable for irrigation and industrial purposes. Signs of deterioration are evident, which can impact marine animals. Ensuring suitable conditions becomes crucial for their survival.
- Very poor (76-100): Limited to irrigation, water quality here is significantly compromised. It poses risks to both humans and marine ecosystems. Special attention is needed to protect vulnerable species.
- Unsuitable (Above 100): Severely degraded water quality endangers marine animals. Remediation efforts are essential to prevent harm to aquatic life.

4) Examination of Class Distribution and Data Augmentation: The class distribution of a dataset refers to the proportion of each class in the dataset. In other words, it is the number of examples of each class divided by the total number of examples in the dataset. Class distribution is an important concept in machine learning because it can affect the performance of the model. If the class distribution is imbalanced, meaning that one class has significantly more examples than the others, then the model may be biased or skewed towards the majority class and perform poorly on the minority class. We studied the class distribution of the water quality classes to address class imbalance and data scarcity problems. This also helped to prepare a well-balanced water quality dataset.
The class distribution of the water quality classes is shown in Figure 5. In Figure 5, class 1 represents "unsuitable," class 2 represents "very poor," class 3 represents "poor," class 4 represents "good," and class 5 represents "excellent." Class 1 has 404 samples, which implies 24% of the entire dataset; class 2 has 328 samples, which represents 19% of the water quality data; class 3 has 668 samples, which represents 40% of the water quality data; class 4 has 265 samples, which implies 15% of the data; and class 2 has three samples, which represents 0.12% of the data.

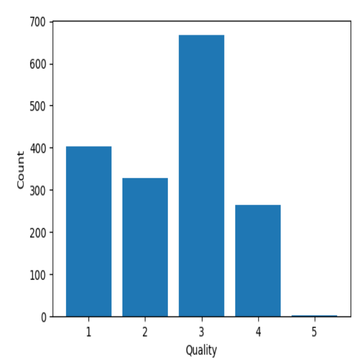


Fig. 5: Class distribution of the water quality classes before data augmentation

The class distribution of the water quality classes in the dataset shown in Figure 5 indicates a class imbalance problem.

This article has been accepted for publication in IEEE Sensors Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2024.3383887

TABLE I: Permissible limits of water quality parameters for computing WQI. [4], [18], [19]

Parameters	Permissible limits
Nitrate (mg/l)	45
pH	8.5
BOD (mg/l)	5
Fecal coliform (Cfu/100 ml)	100
Conductivity (μS/cm)	1000
Temperature (°C)	1
DO (mg/l)	10
Total coliform (Cfu/100 ml)	1000

TABLE II: Categories of WQI and repsective water quality status and use cases. [20]

Range of WQI	Water quality status	Potential use cases
0 - 25	Excellent	Drinking, marine ecosystems, irrigation, and industrial
26 - 50	Good	Drinking, marine ecosystems, irrigation, and irrigation
51 - 75	Poor	Irrigation and industrial
76 - 100	Very poor	Irrigation
Above 100	Unsuitable	Treatment needed before use

Training a water quality classifier with an imbalanced dataset can lead to bias or skewness toward the majority class. Such bias adversely affects model performance, as it may favor the majority class and result in biased classifications. To address this issue, we employed the Synthetic Minority Oversampling Technique (SMOTE) augmentation method [30], [31] from the imbalanced-learn library in Python.

SMOTE enhanced model performance by balancing class distribution and generating synthetic samples for the minority class. By interpolating between existing instances, SMOTE ensured that the model effectively learns from both classes. Additionally, SMOTE mitigated data scarcity by creating synthetic data points, enriching the dataset. A new class distribution is obtained after applying the SMOTE technique, as shown in Figure 6.

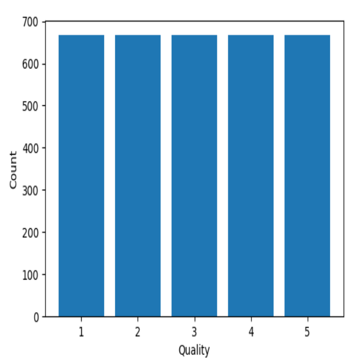


Fig. 6: Class distribution of the water quality classes after data augmentation

5) Data Transformation: This process normalizes the range of features of a dataset to a common range as the range may vary a lot. The variations or differences in the range of features in the data can make the features with a large values to potentially skew or dominate the training speed of the learning algorithm, resulting to a slower convergence in the iterations of learning algorithm, and have a large impact on the classification results. In addition, the problem of skewness in the data in the presence of outliers can affect the learning of the algorithm and make the algorithm to be skewed or biased towards the skewed distributions. To address this, we improved on the data transformation challenges faced by the existing water quality classifiers that used min-max and standard scaler method [12], [13]. For example, the min-max scaler method is affected by outliers. This implies that the presence of outliers can make data scaler inefficient when using the min-max scaler method. To address the problem of the min-max and standard scaler methods, we used the robust scaler method, which is more robust to outliers. The robust scaler method was used to normalize the data to a common range to reduce the impact of different magnitudes on the classification results. The robust scaler method also helped to reduce the skewness in the data in the presence of outliers to reduce the effect skewness on the performance of the learning algorithm.

To handle outliers in the water quality data during the preprocessing, the robust scaler method employed the median and interquartile range (IQR). The IQR calculates the difference between the 75th percentile (Q3) and 25th percentile (Q1) of the water quality data. To normalize the data, the robust scaler method calculates the median and IQR for each water quality feature. Then, it subtracted the median and divided by the IQR using Eqn. 5. By using percentiles, the robust scaler method is robust against extreme values and makes it suitable for the water quality dataset containing outliers.

$$V = \frac{x - median}{Q3 - Q1} \tag{5}$$

where V is the normalized value, x is the original values of each feature, $Q1$ represents the 25th percentile, and Q3 denotes the 75th percentile.

6) Data Partitioning: We employed the three-way holdout method to partition the preprocessed water quality data to ensure rigorous evaluation, unbiased testing, and efficient

This article has been accepted for publication in IEEE Sensors Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2024.3383887

use of available data. The data was partitioned into three sets: training, test, and validation. The partitioning was done according to an 80:10:10 ratio. The training set was utilized for model hyperparameter tuning and training. The validation set allowed us to evaluate the model's performance during hyperparameter tuning and training. Finally, the reserved test set was employed to assess how well the models generalize to new or unseen water quality data. To achieve this partition, we used the train_test_split function in the scikit-learn library.

E. Proposed Highway-BiLSTM Water Quality Classifier

We propose a Highway-BiLSTM method for water quality classification to address the limitations of existing water quality classifiers. The proposed method uses a BiLSTM network [32] and highway network mechanism [33] to process and classify complex and nonlinear water quality data as well as optimize information flow. The proposed classifier consists of the BiLSTM input layer, BiLSTM layer, highway layer, flatten layer, dense layer, and water quality classification layer. The architecture of the proposed Highway-BiLSTM method for water quality classification is shown in Fig. 7.

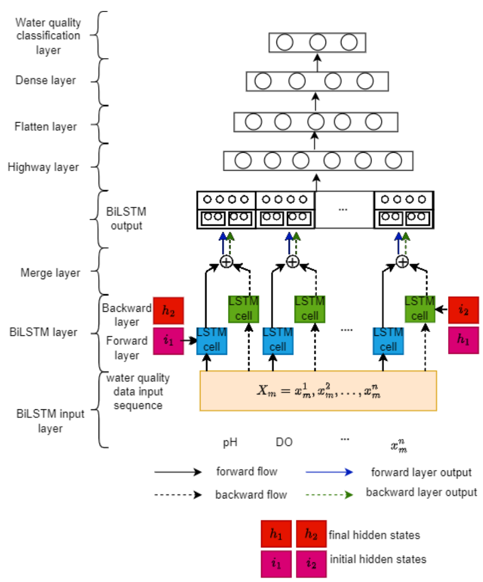


Fig. 7: Architecture of the proposed water quality classification method

The proposed classifier employs an input layer of the BiLSTM network. This layer takes the preprocessed water quality data as its input. The BiLSTM layer of the proposed classifier used two types of LSTM cells: the forward LSTM cell in the forward layer and the backward LSTM cell in the backward layer. The forward LSTM cell processed the input water quality data sequence in the forward direction. It has three gates: the input gate, forget gate, and output gate. The input gate controls what information can be added to the current LSTM cell. The forget gate determines the information to discard from the LSTM cell state and the amount of information the LSTM cell will receive from the LSTM cell from the previous step. The output gate controls the output value of the cell state. The backward LSTM cell processed the water quality data sequence in the backward direction.

In each time step, the forward LSTM cells captured long-term dependencies of the input water quality data $X_m = \{x_m^1, x_m^2, ..., x_m^n\}$ in the forward direction from x_m^1 to x_m^n. Simultaneously, the backward LSTM cells captured long-term dependencies of the water quality data sequence in the backward direction from x_m^n to x_m^1. To create a comprehensive representation after processing the input water quality data sequence, the outputs of the forward and backward LSTM cells in the hidden layers were concatenated. This combined representation captures both past and future context of the input water quality data for each time step. By leveraging information from both directions, the BiLSTM layer effectively captures complex and nonlinear patterns within the water quality data. Furthermore, the concatenated outputs of the entire input water quality data serve as the global feature representation for the entire dataset. The output of the BiLSTM layer was fed into a highway network layer for further processing.

The highway layer in the proposed classifier serves as an information flow optimization scheme for processing and refining features from the BiLSTM layer. This layer introduces two gates: the transformation gate and the carry gate. The transform gate applied the ReLU function to the BiLSTM output (which serves as the input to the highway network). Its purpose is to control the extent to which the input is transformed and passed to the next layer. The carry gate, on the other hand, applies the sigmoid function to the same BiLSTM output. It determines how much information from the input is retained and passed to the next layer. By leveraging these gates, the network achieves better feature refinement, optimizes information flow, and learns which parts of the input are relevant for the task. Consequently, it selectively propagates relevant information while suppressing irrelevant details. Finally, the output of the highway network layer is passed to a flatten layer.

The proposed classifier used the flatten layer to transform the output from the highway network layer into one dimension. This was done to convert the multi-dimensional output of the highway layer into a one-dimensional array that can be passed to the next layer. The one-dimensional data was then passed onto a dense layer.

The dense layer of the proposed classifier was used to learn the complex patterns between the water quality features and the target to obtain a more informative representation. The output of the dense layer was passed to the next layer in the network, which is the water quality feature classification layer.

The water quality feature classification layer of the proposed classifier is a fully connected layer that uses a Softmax activation function to produce the final output of the model for a given water quality data. Each neuron in the layer is activated using the ReLU activation function [34]. The output of this layer is a probability distribution over the different classes of water quality features.

This article has been accepted for publication in IEEE Sensors Journal. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2024.3383887

F. Explainability of the Proposed Water Quality Classifier

We introduced the use of SHAP (SHapley Additive exPlanations) method to address the problem of lack of explainability of the decision making behind the water quality data classification of the existing water quality classifiers which are black box. The problem of lack of explainability makes it challenging to directly understand, visualize and explain the inner workings of prior water quality classifiers.

SHAP is a model-agnostic framework for machine learning model interpretability [35]. It was employed in this study to understand and explain the decision making behind each water quality sample. For this purpose, it provides a visualization that explains how each feature of a water quality sample influenced the classifier's decision. It provides these explanations at the local and global levels by approximating the behavior of the model around specific instances. The local explanations provide insights into the decision making behind a single water quality data. The global explanations helps to understand the general behaviour of the proposed water quality classifier.

The SHAP method uses the Shapley value of each feature in the water quality dataset to calculate the feature importance or contributions of each of the features on the final outcome of the proposed water quality classifier. The Shapley value is a solution concept in cooperative game theory that helps to fairly distribute the total gains among the players in a game. SHAP uses the mathematical model in Eqn. 6 to calculate the Shapley value of feature (player) i.

$$\phi_i(N,v) = \frac{1}{|N|!} \sum_{S \subseteq N \setminus \{i\}} [v(R_i^S \cup \{i\} - v(R_i^S)] \quad (6)$$

where $\phi_i(N,v)$ is the Shapley value for an individual feature i. N represents the number of features. R is the subset of N, and it denotes the permutation of order of the features. The denominator $|N|!$ represents the number of permutations of the set of N features. $v(R \cup \{i\})$ is the contributions made by the set of features, including i, on the outcome for possible permutations of order of the features. $v(R)$ is the contributions made by the set of features, without i, on the outcome for possible permutations of order of the features.

G. Baseline Methods

In this section, we discuss the recent state-of-the-art machine learning methods that are used as baseline methods to compare the performance of the proposed method.

1) Random Forest Method: An RF model for water quality classification was developed using the RandomForestClassifier class in scikit-learn library in Python similar to Khan et al. [16]. RF is an ensemble learning algorithm that combines or generates multiple decision trees to solve classification problems. It achieves this by randomly selecting subsets from the training set to create a collection of decision trees. Each decision tree within the forest independently votes or predicts the output. During testing, these individual votes are aggregated to make the final decision.

2) Gradient Boosting Method: We used the gradient boosting algorithm from the sklearn.ensemble module in the scikit-learn library to develop gradient boosting model for water quality classification similar to Khan et al. [16]. Gradient Boosting is an ensemble learning algorithm that combines multiple decision trees as classifiers to solve classification problems. To achieve this, it begins with a shallow tree to predict the target variable based on features. Subsequently, errors are computed, and subsequent trees are added to correct mistakes made by the previous model. Each new model contributes to classifications, gradually improving performance. The ensemble combines the classifications of these decision trees to enhance accuracy. Additionally, the Gradient Boosting algorithm employs regularization techniques to prevent overfitting.

3) Suppport Vector Machine Method: We utilized the sklearn.svm module from the scikit-learn library to develop the SVM model for water quality classification similar to Ladjal et al. [13]. An SVM is a machine learning algorithm commonly employed for solving classification tasks. The SVM method divides the data samples (support vectors) of a dataset into distinct classes using hyperplanes (lines). Subsequently, the best hyperplane that accurately separates the classes is selected. This optimal hyperplane corresponds to the line with the largest margin, where the margin represents the gap between the hyperplanes on the nearest data samples of different classes. SVM models are typically constructed using various kernel functions, such as the sigmoid, radial basis, and polynomial kernels. To determine the most efficient kernel function for water quality classification, we employed an optimization method.

4) ANN Method: We used the tensorflow.keras library to develop the ANN model for water quality classification similar to Ladjal et al. [13]. ANN is a machine learning algorithm that is used for solving classification problems. It was inspired by the way nerve cells function in the human brain. It consists of interconnected layers, with information flowing unidirectionally from the input layer to the output layer. These densely connected layers adaptively transform data through a series of hidden units, allowing the ANN to understand complex patterns. ANNs use learning algorithms to adjust their output based on errors during training, ultimately minimizing differences between predicted and actual outcomes.

H. Evaluation Metrics

To evaluate the classification performance of the proposed BiLSTM method and the baseline methods, standard classification metrics were considered. These included accuracy, precision, specificity, sensitivity, F1-score, and confusion matrix. The accuracy is a metric that evaluates how well a model can make correct classifications. It is calculated by dividing the number of correct classifications by the total number of classifications. A higher accuracy means that the model has a

This article has been accepted for publication in IEEE Sensors Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2024.3383887

lower error rate. The accuracy of a model can be calculated using Eqn. 7.

$$Accuracy = \frac{TN + TP}{TP + FP + TN + FN} \tag{7}$$

where TP, TN, FN, and FP are the abbreviations for the four possible outcomes of a classification model. TP stands for true positive, meaning that the model correctly predicted a positive value when the actual value was also positive. TN stands for true negative, which means that the model correctly predicted a negative value when the actual value was negative. FN stands for false negative, which means that the model incorrectly predicted a negative value when the actual value was positive. FP stands for false positive, meaning that the model incorrectly predicted a positive value when the actual value was negative.

Precision measures the fraction of positive classifications that are correct, and reflects how well a model can avoid false positives. The precision of a models can be calculated using Eqn. 8:

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

Sensitivity, also known as recall, is a metric that assesses how well a model can identify the positive samples correctly. It is calculated by dividing the number of true positives by the number of actual positives. A higher sensitivity means that the model has a lower false negative rate. The sensitivity of a model can be calculated using Eqn. 9:

$$Sensitivity = \frac{TP}{TP + FN} \tag{9}$$

F1-score is a metric that measures the harmonic mean of precision and sensitivity. A higher F1-score means that the model has a better balance of precision and sensitivity. The F1-score of a model can be calculated using Eqn. 10.

$$F1 - score = 2 \times \frac{precision \times sensitivity}{precision + sensitivity} \tag{10}$$

Confusion matrix is a performance metric that evaluates how well a classification model can make correct classifications. It is calculated by organizing the model's classifications into the four categories (TP, TN, FP, FN). The confusion matrix provides valuable insights into the model's strengths and weaknesses to assess its overall performance.

IV. Results and Discussion

This section outlines the experimental environment, describes the approach used for hyperparameter optimization across all models, explains the model training process, and presents the results from various experiments.

A. Experimental Setup

Experiments were performed to compare our proposed method with the baseline methods. The experiments were conducted in Python environment on a computer with the following setup - Operating System: Windows 10 with a 64-bit, Processor: Intel i7 CPU Intel(R) Core(TM) i5-1135G7 CPU @ 2.40 GHz-2.42 GHz, and RAM: 8 GB.

B. Parameter Settings

The random search optimization method was employed from scikit-learn and keras tuner libraries to optimize and select the best hyperparameter values for both the proposed model and the baseline models. We chose the random search optimization method due to its computational efficiency.

The hyperparameter optimization procedure is discussed as follows. During the hyperparameter optimization of the models, the models were trained and evaluated on the training data and validation data, respectively. The validation data helped to select the best hyperparameter values for the models.

We fine-tuned the hyperparameters of the RF model. The search space for the n_estimators hyperparameter included integer values between 10 and 100 (inclusive). For the criterion hyperparameter, the search space consisted of two options: "gini" and "entropy". After optimization, we selected an optimal value of 93 for n_estimators, and the best choice for the criterion was "gini".

We fine-tuned the hyperparameters of the Gradient Boosting model. The search space for the learning rate was uniform values within the range of 0.01 and 0.5. For the n_estimators hyperparameter, we considered integer values between 10 and 100 (inclusive). After optimization, we selected an optimal learning rate of 0.32, and 87 for n_estimators.

We fine-tuned the hyperparameters of the SVM model. The search space for the C hyperparameter included uniform values between 0.1 and 100. For the kernel, we considered options such as "linear", "poly", "rbf", and "sigmoid". Additionally, the search space for the gamma hyperparameter included choices like "scale", "auto", 0.1, 0.5, and 1. After optimization, we selected the following values - kernel: RBF, gamma: 0.1, and C: 10. These hyperparameter values were determined to be the most effective for the SVM model.

We fine-tuned the hyperparameters of the ANN model. The search space for the hidden layers' neurons included integer values between 8 and 64. For the learning rate, we considered uniform values between 0.001 and 0.1. After optimization, we selected the following values - first hidden layer neurons: 64, second hidden layer neurons: 32, third hidden layer neurons: 16, fourth hidden layer neurons: 8, and optimal learning rate: 0.001.

To determine the optimal model architecture for the proposed classifier, we fine-tuned the hyperparameters of the proposed classifier. The search space for the BiLSTM cells included integer values between 16 and 128 with an increment of 16. For the dense layer, we considered integer values between 16 and 128 with an increment of 16 for the neurons as well as "relu" and "tanh" for the activation. The search space for the epochs included {50, 100, 150, 200}. For the batch size, we

This article has been accepted for publication in IEEE Sensors Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2024.3383887

considered a search space of {16, 32, 64}. After optimization, we selected the best model architecture with one BiLSTM layer, 16 LSTM cells in the BiLSTM layer, 16 neurons and tanh activation function in the dense layer, and a batch size of 64. In addition, the best model architecture selected through the hyperparameter procedure helped to ensure that the performance of the model is robust across different parameter settings.

C. Model Training

During the training process, the best model selected using a random search optimization and three-way holdout procedure was trained using the training data and validation data. The model was compiled using the Adam optimizer with a learning rate of 0.001, a categorical cross-entropy loss function, and an accuracy metric. It was then trained on the input training data for an epoch of 150 with a batch size of 64. The model parameters were updated using the Adam optimization algorithm to minimize the categorical cross-entropy loss function. The optimizer adjusted the weights and biases of the model to reduce the error. The validation data was used to evaluate the performance of the model during training.

D. Analysis of the Proposed Model Performance

In this section, we studied the performance and effectiveness of the proposed method using the training and learning curves. The training curves were used to evaluate the performance of the proposed model during the training process. As presented in Figure 8, the training curves plot the training and validation accuracies of the proposed model against the number of epochs (i.e., the number of training iterations). The training accuracy measured how well the proposed model performed on the training dataset, while the validation accuracy measured its performance on a separate validation dataset that the model had not seen during training. From Figure 8, it was observed that the proposed method achieved a better performance in terms of accuracy with a fast convergence speed in training and validation. This was deduced from the validation accuracy curve, which is very similar to the training accuracy curve.

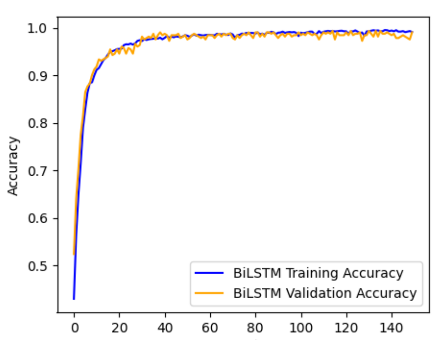


Fig. 8: Training and validation accuracy across epochs

Next, we evaluated the performance of the proposed model on the training and validation datasets after each update using the learning curves. Learning curves are graphical representations of the learning performance of a model over time. Figure 9 shows the training and validation losses of the proposed method, respectively. From Figure 9, the training loss curve shows that the proposed method performs well in learning the water quality data over time, despite the complex characteristics of the data. In addition, the validation loss curve shows that the proposed method has good generalization ability.

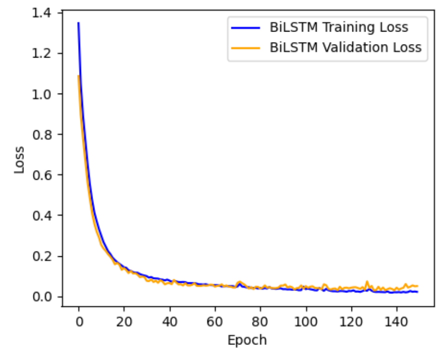


Fig. 9: Training and validation loss across epochs

E. Performance Evaluation and Validation

To evaluate the generalizability of the proposed classifier in handling unseen data, we assessed its performance using previously unseen water quality data. We employed several metrics to compare its classifications against the actual values. Additionally, we validated its robustness by comparing its performance to that of baseline classifiers. Table III presents the evaluation results of the proposed and baseline methods.

As shown in Table III, the proposed Highway-BiLSTM outperformed the baseline water quality classifiers in classifying water quality features in terms of accuracy, precision, sensitivity, and F1-score.

The proposed method achieved a 2% increase in accuracy, precision, sensitivity, and F1-score for classifying water quality compared to the RF method. Furthermore, it outperformed the Gradient Boosting method by 4% in accuracy, precision, sensitivity, and F1-score on water quality classification tasks. In addition, when classifying water quality, it surpassed the SVM method by 4% in accuracy, sensitivity, and F1-score. The proposed method demonstrated a 3% increase in precision compared to the SVM method. Similarly, the proposed method also surpassed the ANN method on classifying water quality with 1% accuracy, precision, sensitivity, and F1-score.

The improvements of the proposed method are a result of its ability to handle complex data. It achieved this by effectively managing intricate data, which allows it to capture long-distance dependencies within water quality sequences. Secondly, it leveraged bidirectional context from water quality features to extract more relevant information.Thirdly, it benefited from the high quality water quality data obtained by the improved data preprocessing decision introduced in this work. Examples include the use of SMOTE technique that was employed to address class imbalance and data scarcity

This article has been accepted for publication in IEEE Sensors Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2024.3383887

TABLE III: Performance assessment and comparison of different models on unseen water quality data

Model	Accuracy	Precision	Sensitivity	F1-score	Training Time (s)	Classification Time (s)
RF	96	96	96	96	0.482	0.014
Gradient Boosting	94	94	94	94	2.572	0.004
SVM	94	95	94	94	0.108	0.017
ANN	97	97	97	97	0.564	0.069
Highway-BiLSTM	**98**	**98**	**98**	**98**	0.407	0.077

problems. The water quality data we obtained by applying robust scaler ensures that the proposed model is less sensitive to outliers. Thirdly, the hyperparameter optimization procedure helped to select optimal hyperparameter values that further enhances the effectiveness of the proposed method.

A comparison of the models in Table III indicates that the proposed method is superior to the baseline methods and more efficient in predicting water quality.

Furthermore, it is evident from Table III that the proposed method outperformed simple machine learning algorithms such as RF, Gradient Boosting, and ANN in terms of training time. This achievement is remarkable, especially considering the computational time constraints imposed by the BiLSTM algorithm. Additionally, the proposed method demonstrated exceptional efficiency in classifying water quality samples within the test set of 334 samples, achieving this task in less than 1 second. The success of the proposed method in both training time and classification time can be attributed to its utilization of a highway network mechanism. This mechanism optimizes information flow within the network, facilitating efficient computations and accurate classifications.

F. Model Interpretability

We used the KernelExplainer from the SHAP library in Python to systematically investigate and determine the features of water quality data that contributed to or impacted the classifications of the proposed model. This helped to interpret the classifications of the proposed method. To achieve this purpose, we randomly selected water quality samples from the training set and the test set.

To elucidate the classifications of the proposed method, we initially constructed a SHAP explainer (which acts as a surrogate model) by leveraging the proposed model and the training dataset. Specifically, we employed the training set samples as background data to train this surrogate model. The model was built using various water quality features, and we utilized the KernelExplainer function for this purpose.

To provide a comprehensive understanding of the decision-making process behind the proposed water quality classifier, we employed the created SHAP explainer. First, at the local level, we calculated the SHAP values for an individual water quality data point randomly selected from the test sample. This localized explanation helped us assess the contribution (importance) of each feature in the model's decision. Additionally, for global interpretability, we computed the SHAP values for all test samples and summed the absolute SHAP values for each individual water quality data classification. These aggregated values provided insights into how individual features impact the model's classifications, thereby illuminating the underlying mechanisms of our proposed method.

We employed the force plot to visualize how the features of a water quality sample contributed to the classification of the model for a specific water quality sample. Figures 10, 11, 12, and 13 shows the force plot of the SHAP local interpretation results that explains the impact of water quality features as forces on the proposed model's classifications. As can be seen in Figures 10, 11, 12, and 13, each feature value represents a force that shows the contribution of each feature to pushing the classification of the proposed model from the base value to the output $f(x)$ of the model for a specific water quality sample. The base value for the SHAP values is the average of all classifications. Figures 10, 11, 12, and 13 explanations showed that the base values for samples one, two, three, and four are 0.204, 0.207, 0.199, and 0.195.

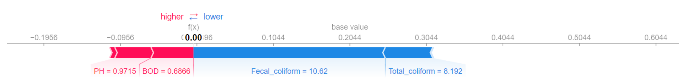


Fig. 10: SHAP local interpretation for the proposed model classifications for sample one

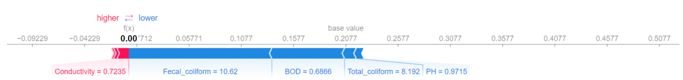


Fig. 11: SHAP local interpretation for the proposed model classifications for sample two

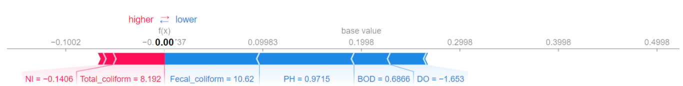


Fig. 12: SHAP local interpretation for the proposed model classifications for sample three

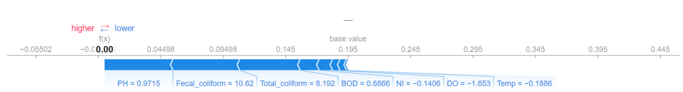


Fig. 13: SHAP local interpretation for the proposed model classifications for sample four

The features that were important for making the classification for each water quality sample are shown in blue and red, respectively. The features pushing the model's classification lower are shown in red, whereas the features shown in red contribute to increasing the classification of the model. Moreover, the features that had a greater impact on the output of

This article has been accepted for publication in IEEE Sensors Journal. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2024.3383887

the model were located close to the boundary between blue and red. The magnitude of the impact was determined by the size of the bar.

Figure 10 explanation showed that pH and BOD were the features that increased the model's classification higher, while fecal coliform and total coliform were the features that pushed the classification of the proposed model lower towards $f(x)$ for sample one.

Figure 11 explanation showed that one feature increased the model's classification towards $f(x)$ for sample two. Figure 12 and 13 explanations also indicated the features that impacted the classification of the proposed model for samples three and four.

Figure 14 explanation showed the global visualization results of the proposed model. This indicated that pH was the most important feature, followed by BOD, fecal coliforms, and DO. Moreover, it shows the impact of each feature on the water quality classes. For instance, the pH feature has a large impact on the excellent water quality class. The BOD feature had a large impact on the poor and very poor water quality classes. Fecal coliform feature had a small impact on the excellent water quality class. In addition, the nitrate and temperature features had the least impact on the water quality class, which was not suitable for drinking.

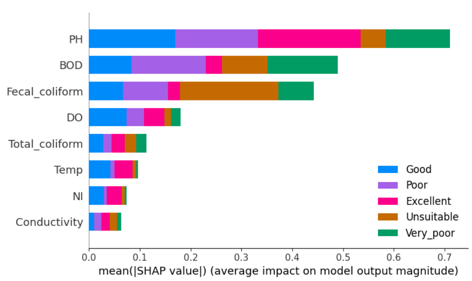


Fig. 14: SHAP global interpretations for the proposed model classifications

How transferable are the insights gained from the SHAP method to real-world applications of the water quality classifier, and how can they be utilized to improve decision-making processes in environmental management?

The insights gained from the SHAP method hold significant value for real-life and practical applications. For instance, SHAP values play a crucial role in decision-making by helping us understand why specific classifications were made. This transparency is essential for decision-makers, particularly in environmental management contexts. SHAP insights also aid in detecting changes in water quality. By identifying critical features that impact water quality, we can assess risks and guide resource management to prevent environmental challenges, such as biodiversity loss. Moreover, these insights empower decision-makers, contributing to informed environmental management.

Practically, decision-makers can utilize SHAP insights to formulate effective policies. For example: if pH levels significantly impact water quality, policies can target pH adjustments to neutralize acidic or alkaline water. When microbiological indicators (such as faecal and total coliforms) significantly affect water quality, policies can focus on reducing their levels. For nitrate levels that significantly impact water quality, policies can aim at nutrient reduction. In cases where dissolved oxygen levels significantly affect water quality (e.g., low dissolved oxygen indicating water pollution or excessive nutrient loading), policies can address increasing dissolved oxygen levels to preserve biodiversity and protect aquatic ecosystems. Additionally, SHAP can guide the development of early warning systems by detecting deviations from expected water quality based on feature contributions. These SHAP explanations can inform adaptive management strategies, allowing interventions to be adjusted based on real-time data and feature importance.

G. Error Analysis

We further emphasized the superiority of our proposed method by comparing it with the baseline methods. In the confusion matrix results for both the proposed method and the baseline methods (as depicted in Figs. 15, 16, 17, 18, and 19), the diagonal represents the percentage of correctly classified water quality samples. However, it is crucial to note that the off-diagonal elements correspond to incorrectly classified samples. Upon analyzing the confusion matrices, we observed that our proposed method achieved high accuracy across all five classes. A notable trend in the misclassification patterns between the proposed water quality classifier and the baseline classifiers' confusion matrices is that the baseline water quality classifiers consistently struggle with the "poor" and "very poor" water quality data categories. In contrast, the proposed classifier consistently classified 'very poor' and 'poor' water quality samples. This observed trend may be attributed to the baseline classifiers' limited ability to capture the complex and nonlinear patterns within the water quality data. Furthermore, it demonstrated rare errors in classifying water quality samples in the 'very poor,' 'poor,' and 'good' classes, outperforming the Gradient Boosting classifier.

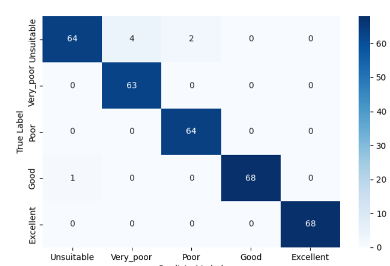


Fig. 15: Confusion matrix for the proposed method

V. Conclusion

In this study, we introduced a Highway-BiLSTM-based water quality classifier designed for an edge-enabled water quality monitoring system that performs on-site water quality classifications. To ensure reliability, we addressed data preprocessing challenges commonly associated with water

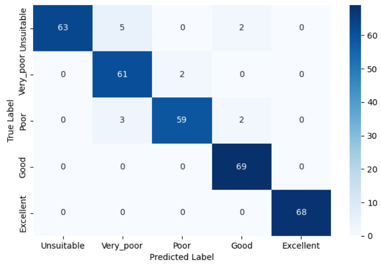


Fig. 16: Confusion matrix for the RF classifier

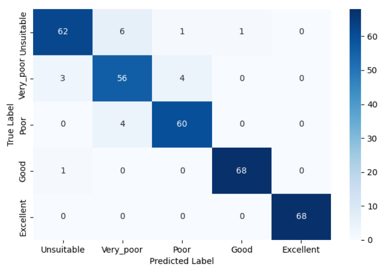


Fig. 17: Confusion matrix for the Gradient Boosting classifier

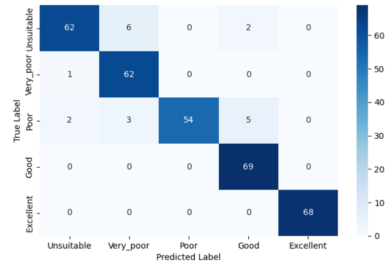


Fig. 18: Confusion matrix for the SVM classifier

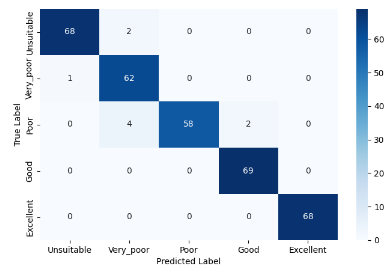


Fig. 19: Confusion matrix for the ANN classifier

quality classifiers in existing literature. Our approach involved using the SMOTE technique to address the class imbalance and water quality data scarcity problems. We employed a robust scalar normalization method, leveraging advanced statistical techniques (such as median and interquartile range), to reduce skewness and address outliers during data preprocessing. The proposed water quality classifier utilizes a BiLSTM network. BiLSTM captures bidirectional context from water quality features, extracting crucial information and learning complex, nonlinear patterns within the data. Several parameters such as fecal coliform, total coliform, pH, biological oxygen demand, dissolved oxygen, conductivity, nitrate, and temperature levels contribute to the classification process. We incorporated a highway network mechanism to enhance information flow within the model. The proposed Highway-BiLSTM-based classifier outperformed baseline water quality classifiers across various standard metrics, including accuracy, precision, sensitivity, F1-score, and confusion matrix. This robust performance confirms the classifier's reliability and its ability to generalize well to unseen water quality data. In practical terms, the deployment of the proposed edge-enabled water quality monitoring system holds significant promise. For example, it bridges the gap between data collection and timely actionable insights. Its implementation promises a more resilient, informed, and sustainable approach to managing our invaluable water resources. This can help to safeguard both public health and the environment. Despite the success of the proposed method, deep learning algorithms such as BiLSTM require large training samples. This limitation will be addressed further in future research by collecting more water quality data training samples.

References

[1] S.O. Olatinwo, and T. H. Joubert, "Energy efficient solutions in wireless sensor systems for water quality monitoring: A reiew," *IEEE Sensors Journal*, vol. 19, no. 5, pp. 159 - 1625, 2019.

[2] N. Radhakrishnan, and A.S. Pillai, "Comparison of water quality classification models using machine learning," in *Proc. 2012 International Conference on Communication and Electronics Systems (ICCES)* IEEE, 2020, pp. 1183-1188.

[3] S.O. Olatinwo, and T. H. Joubert, "Deep learning for resource management in Internet of Things networks: A bibliometric analysis and comprehensive review," *IEE Access*, vol. 10, no. 5, pp. 94691 - 94717, 2022.

[4] WHO. *Guidelines for drinking-water quality 2017*. [Online]. Available: https://www.who.int/publications/i/item/9789241549950. Accessed on: May 19, 2022.

[5] South African Water Sanitation. *No Drop Watch Report 2023*. [Online]. Available: https://www.sanews.gov.za/south-africa/report-shows-decline-drinking-water-quality. Accessed on: Mar. 11, 2023.

[6] UNESCO. *Imminent risk of a global water crisis, warns the UN World Water Development Report 2023*. [Online]. Available: https://www.unesco.org/en/articles/imminent-risk-global-water-crisis-warns-un-world-water-development-report-2023. Accessed on: Mar. 11, 2023.

[7] UNEP. *Globally, 3 billion people at health risk due to scarce data on water quality*. [Online]. Available: https://www.unep.org/news-and-stories/story/globally-3-billion-people-health-risk-due-scarce-data-water-quality. Accessed on: Mar. 11, 2023.

[8] S.O. Olatinwo, and T. H. Joubert, "A bibliometric analysis and review of resource management in internet of water things: The use of game theory," *Water*, vol. 14, p. 1636, 2022.

[9] J. Bi, Z. Chen, H. Yuan, and J. Zhang, "Accurate water quality prediction with attention-based bidirectional lstm and encoder-decoder," *Expert Systems with Applications*, p. 121807, 2023.

[10] M. Miller, A. Kisiel, D. Cembrowska-Lech, I. Durlik, and T. Miller, "IoT in water quality monitoring—are we really here?," *Sensors*, vol. 23, p. 960, 2023.

[11] H. Chen, J. Yang, X. Fu, Q. Zheng, et al., "Water quality prediction based on lstm and attention mechanism: A case study of the burnett river, Australia," *Sustainability*, vol. 14, p. 13231, 2022.

[12] P. Liu, J. Wang, A.K. Sangaiah, Y. Xie, and X. Yin, "Analysis and prediction of water quality using LSTM deep neural networks in IoT environment," *Sustainability*, vol. 11, p. 2058, 2019.

[13] M. Ladjal, M. Bouamar, Y. Brik, and M. Djerioui, "A decision fusion method based on classification models for water quality monitoring," *Environmental Science and Pollution Research*, vol. 30, pp. 22532-22549, 2023.

[14] M.I. Khelil, M. Ladjal, Y. Brik, and M.A. Ouali, "Self-organizing maps-based features selection with deep LSTM and SVM classification approaches for advanced water quality monitoring," *International Journal of Intelligent Enginnering & Systems*, vol. 15, pp. 90-102, 2022.

[15] S. Dilmi, and M. Ladjal, "A novel approach for water quality classification based on the integration of deep learning and feature extraction techniques," *Chemometrics and Intelligent Laboratory Systems*, vol. 214, p. 104329, 2021.

[16] S.Y. Abuzir, and Y.S. Abuzir, "Machine learning for water quality classification," *Water Quality Research Journal*, vol. 57, pp. 152-164, 2022.

[17] M.S.I. Khan, N. Islam, J. Uddin, S. Islam, and M.K. Nasir, "Water quality prediction and classification based on principal component regression and gradient boosting classifier approach," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, pp. 4773–4781, 2022.

[18] DWAE. *South African water quality guidelines*. [Online]. Available: https://www.iwa-network.org/filemanager-uploads/WQ_Compendium/Database/Selected_guidelines/041.pdf. Accessed on: May. 10, 2022.

[19] WHO. *Manganese in drinking-water: background document for development of WHO guidelines for drinking-water quality*, Technical documents.[Online]. Available: https://www.who.int/docs/default-source/wash-documents/wash-chemicals/manganese-background-document.pdf. Accessed on: May. 11, 2022.

[20] S. Tyagi, B. Sharma, P. Singh, and R. Dobhal, "Water quality assessment in terms of water quality index," *American Journal of Water Resources*, vol. 1, pp. 34-38, 2013.

[21] D.D Olatinwo, A.-M. Adnan, H. Gerhard, and M. Hermanus, "IoT-enabled WBAN and machine learning for speech emotion recognition in patients," *Sensors*, vol. 23, no. 6, pp. 1-23, 2023.

[22] D.D Olatinwo, A.-M. Adnan, and H. Gerhard, "Energy efficient priority-based hybrid MAC protocol for IoT enabled WBAN systems," *IEEE Sensors Journal*, vol. 23, no. 12, pp. 13524-13538, 2023.

[23] D.D Olatinwo, A.-M. Adnan, and H. Gerhard, "Energy-aware hybrid MAC protocol for IoT enabled WBAN systems," *IEEE Sensors Journal*, vol. 22, no. 3, pp. 2685-2699, 2021.

[24] Kaggle. *Indian water quality data*. [Online]. Available: https://www.kaggle.com/datasets/anbarivan/indian-water-quality-data. Accessed on: Jan. 20, 2022.

[25] K. Kumar, and B. Pande, B., "Air pollution prediction with machine learning: A case study of Indian cities," *International Journal of Environmental Science and Technology*, vol. 20, no. 5, pp. 1-6, 2022.

[26] C.D. Muñiz, P.G. Nieto, J.A. Fernández, J.M. Torres, and J. Taboada, "Detection of outliers in water quality monitoring samples using functional data analysis in San Esteban estuary (Northern Spain)," *Science of the Total Environment*, vol. 439, pp. 54-61, 2012.

[27] I. Rangeti, B. Dzwairo, G.J. Barratt, and F. Otieno, "Validity and errors in water quality data-a review," *Research and Practices in Water Quality, Durban University of Technology, Durban, South Africa*, pp. 92-112, 2015.

[28] P.D. Talagala, R.J. Hyndman, C. Leigh, K. Mengersen, and K. Smith-Miles, "A feature-based framework for detecting technical outliers in water-quality data from in situ sensors," *Water Resources Research*, vol. 55, pp. 8547–8568, 2019.

[29] G. Carrasco, J.L. Molina, M. Patino-Alonso, M.D.C. Castillo, M.P. Vicente-Galindo, and M.P. Galindo-Villardón, "Water quality evaluation through a multivariate statistical HJ-Biplot approach," *Journal of Hydrology*, vol. 577, p. 123993, 2019.

[30] M. Buda, A. Maki, and M.A. Mazurowski, "A systematic study of the class imbalance problem in convolutional neural networks," *Neural Networks*, vol. 106, pp. 249–259, 2018.

[31] R. Mohammed, J. Rawashdeh, and M. Abdullah, "Machine learning with oversampling and undersampling techniques: Overview study and experimental results," in *Proc. 2020 11th international conference on information and communication systems (ICICS)* IEEE, 2020, pp. 243–248.

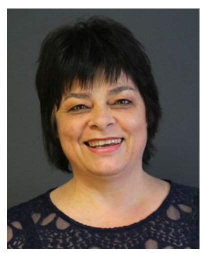
[32] S. Hochreiter, and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 1997.

[33] J.G. Zilly, R.K. Srivastava, J. Koutnık, and J. Schmidhuber, "Recurrent highway networks," in *Proc. PMLR 2017 international conference on machine learning*, 2017, pp. 4189-4198.

[34] A.F. Agarap, "Deep learning using rectified linear units (ReLU)," *arXiv preprint*, arXiv:1803.08375, 2018.

[35] S.M. Lundberg and S.I. Lee, "A unified approach to interpreting model predictions," in *Proc. 2017 31st Conf. Neural Information Processing Systems (NIPS 2017)*, CA, USA, NIPS, 2017, pp. 1–10.

Segun O. Olatinwo is a researcher with research interests in artificial intelligence, environmental monitoring, marine biodiversity preservation, public health protection, agricultural productivity, big data analytics, model interpretability, computational efficient models for resource constrained systems, educational technology, intelligent systems / smart networks, Internet of Things / sensors, signal processing, computing technologies, energy harvesting, resource management, and quality of service performance.



Trudi-H. Joubert (Senior Member, IEEE) received the B.Eng. and M.Eng. degrees (Hons.) in electronic engineering and the Ph.D. degree from the University of Pretoria, South Africa. She is currently the Director of the Carl and Emily Fuchs Institute for Microelectronics (CEFIM) and an Assistant Professor with the Department of Electrical, Electronic and Computer Engineering, University of Pretoria. Her core pursuit is CMOS and BiCMOS mixed signal circuit design, mainly in readout and signal processing for imaging arrays, biologically inspired electronic circuit applications, and specialized devices for analog IC design. Her most recent endeavor is printed electronics sensors and circuits, particularly for water quality monitoring, but also in point-of-use diagnostics. She is an inventor on three patents, has published 22 journal articles, and presented and coauthored 65 peer-reviewed conference proceedings papers. She has lectured short courses to industry in her field of specialization, worked on 31 projects for the South African industry, writing 131 technical reports in the process, and contributing to 32 integrated circuit chips. The contributions of her work to the South African innovation system spans university teaching, applied research, and leading industrial development of prototypes and products. She is registered as a Professional Engineer with the Engineering Council of South Africa (ECSA). She has been a member of both the technical and the organizing committee for several international conference events.

PLACE PHOTO HERE

Damilola D. Olatinwo is a researcher and an academician who has research expertise in the mathematical modelling of problems in research areas like internet of things, wireless body area networks, healthcare monitoring, machine-to-machine communications, low power and long-range communication systems. Her research expertise also include the design of resource management strategies, communication protocols, and the application of artificial intelligence techniques, development of stochastic probability models.