Predicting Urban Water Consumption and Health Using Artificial Intelligence Techniques in Tanganyika Lake, East Africa

Alain Niyongabo 1,2, Danrong Zhang 1,4,*, Yiqing Guan 1, Ziyuan Wang 1, Muhammad Imran 1, Bertrand Nicayenzi 5, Alemayehu Kbeta Guyasa 1 and Pascal Hatungimana 3

1 College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China; lx20220601006@hhu.edu.cn (A.N.); yiqingguan@hhu.edu.cn (Y.G.); wang_ziyuan@hhu.edu.cn (Z.W.); m.imrankbr@gmail.com (M.I.); gkabeta19@gmail.com (A.K.G.)
2 Burundian Agency of Rural Hydraulics and Sanitation (AHAMR), RN2, Gitega P.O. Box 176, Burundi
3 Faculty of Engineering Sciences, University of Burundi, Belvedere Road, Bujumbura P.O. Box 2720, Burundi; pascalhatungimana@gmail.com
4 College of Geography and Remote Sensing, Hohai University, Nanjing 210028, China
5 College of Civil Engineering and Transportation, Hohai University, Nanjing 210098, China; bnicayenzi1@gmail.com

* Correspondence: danrong_zhang@hhu.edu.cn

Abstract: Water quality has significantly declined over the past few decades due to high industrial rates, rapid urbanization, anthropogenic activities, and inappropriate rubbish disposal in Lake Tanganyika. Consequently, forecasting water quantity and quality is crucial for ensuring sustainable water resource management, which supports agricultural, industrial, and domestic needs while safeguarding ecosystems. The models were assessed using important statistical variables, a dataset comprising six relevant parameters, and water use records. The database contained electrical conductivity, pH, dissolved oxygen, nitrate, phosphates, suspended solids, water temperature, water consumption records, and an appropriate date. Furthermore, Random Forest, K-nearest Neighbor, and Support Vector Machine are the three machine learning methodologies employed for water quality categorization forecasting. Three recurrent neural networks, namely long short-term memory, bidirectional long short-term memory, and the gated recurrent unit, have been specifically designed to predict urban water consumption and water quality index. The water quality classification produced by the Random Forest forecast had the highest accuracy of 99.89%. The GRU model fared better than the LSTM and BiLSTM models with values of R² and NSE, which are 0.81 and 0.720 for water consumption and 0.78 and 0.759 for water quality index, in the prediction results. The outcomes showed how reliable Random Forest was in classifying water quality forecasts and how reliable gated recurrent units were in predicting water quality indices and water demand. It is worth noting that accurate predictions of water quantity and quality are essential for sustainable resource management, public health protection, and ecological preservation. Such promising research could significantly enhance urban water demand planning and water resource management.

Keywords: water quality and consumption forecasting; classification forecast; Lake Tanganyika; deep learning; machine learning; East Africa

1. Introduction

Water is an important natural resource that has economic and social significance for people. The survival of humans would be in jeopardy without water [1]. Surface water and groundwater are the two most significant drinking supplies globally. Currently, over 1.1 billion people worldwide lack access to safe drinking water [2]. Both point and
non-point pollution sources contribute to deteriorating surface water quality (WQ) worldwide [3]. Growing global concern surrounds water quality degradation due to extensive human activity [3]. Developing nations focus on water supply and cleanliness, while industrialized countries prioritize public health and population growth [4,5]. Water quality directly impacts human health, biodiversity, and various uses in the world, including in agriculture and industry, including in Africa [6]. Contaminated water sources pose significant health risks, causing millions of illnesses and fatalities annually, particularly in developing regions, including East Africa [7]. In the same way, the lack of safe drinking water is observed in developing countries, including Africa and East Africa, where Lake Tanganyika is located. One of the most basic requirements for survival is access to water in an appropriate amount and quality.

In previous times, models such as conventional modeling methods, which are based on Box-Jenkins and autoregressive models, were used to assess WQ and water consumption but had several limitations. They often require gathering numerous parameters, relying on previous knowledge and calibration, which can be resource-intensive and hinder their applicability [8]. Similarly, physical-based models necessitate making multiple assumptions and demanding a comprehensive understanding of the subject matter beforehand [9]. For example, a physical model developed in Malaysia required substantial data collection and incurred significant expenses to evaluate floodplain dynamics and water levels [10]. Numerical models emerged as alternatives to address the limitations of physical models, as exemplified by the Yangtze River water level forecasting model by Wu et al. [11]. However, studies like Guan et al. [12] revealed shortcomings in numerical modeling approaches. Despite advancements, numerical models still struggle with replicating specific physical processes accurately. In recent years, data-driven approaches have gained traction for overcoming traditional models’ shortcomings. Machine learning (ML) algorithms, such as radial basis function (RBF) and support vector machines (SVM), have shown promise in forecasting hydrological characteristics like reservoir evaporation [13]. Similar success has been observed in predicting subsurface evaporation rates and daily water levels in reservoirs using SVM [14]. Though some challenges persist with ML algorithms, previous research indicates their superiority over traditional methods in accuracy. Adjusting hyperparameters, like weights and activation functions, is crucial to enhancing ML model performance [15]. Advancements in artificial intelligence (AI), particularly deep learning methods such as long short-term memory (LSTM) and nonlinear auto-regression neural networks (NARNET), have further improved water quality prediction accuracy. In addition, Theyazn and colleagues employed Naïve Bayes (NB), k-nearest neighbors (KNN), and support vector machines (SVM) models to classify water quality. They achieved an accuracy rate of 97.01% using the SVM machine learning technique, and the $R^2$ of NARNET is 96.17% [16]. Additionally, machine learning techniques like decision trees (DT) and boosted decision trees (BDT) have demonstrated success in predicting daily precipitation [17]. Machine learning and deep learning algorithms have been employed in various studies to forecast and categorize water quality, highlighting their versatility and effectiveness [18].

Nonetheless, urban water distribution networks are crucial infrastructure in cities, necessitating intelligent management to ensure sufficient water supply at the desired pressure and quantity [19]. Forecasting short-term water demand is a key aspect of water distribution network operation and management, influenced by factors like temperature, population, and water pressure [20]. Water demand forecasting methods encompass both linear and nonlinear approaches [21]. Linear methods, including exponential smoothing and autoregressive integrated moving averages (ARIMA), rely on time series analysis [22]. However, nonlinear methods such as artificial neural networks (ANNs) tend to offer better accuracy for short-term forecasting. For instance, Ghiassi et al. [23] employed ANNs for water demand prediction. Herrera utilized the support vector machine method. Additionally, fuzzy logic was applied by [24]. Various other ANNs were
employed in previous research for urban water demand prediction, including the generalized regression neural network, radial basis function networks, feedforward neural networks [25], and the extreme learning machine method [26]. Despite their effectiveness, ML models face challenges in feature selection and overfitting [27]. Deep learning methods, like long short-term memory (LSTM) and gated recurrent units (GRU), show promise in improving accuracy for water demand prediction [28]. Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are also employed, with recent studies favoring hybrid DNNs combining both architectures [28–30]. Hybrid DNNs integrate CNNs for spatial feature extraction and RNNs for temporal feature modeling. These models have applications across various domains, including human activity recognition and energy forecasting. Despite the effectiveness of hybrid DNNs, they remain underevaluated in water demand prediction, a task inherently challenging due to the complex time series nature of water demand data [30].

However, natural elements and anthropogenic activities such as mining, urbanization, stormwater, construction waste, non-functional wastewater treatment plants (WWTPs), and agriculture are key factors influencing the WQ of Lake Tanganyika [31]. Point sources, like industrial sites and WWTPs, directly discharge pollutants into the water of Lake Tanganyika, including heavy metals and organic compounds, significantly impacting ecosystems [32]. Millions of people who rely on the resources of this lake for their livelihoods depend on clean water. The water quality of Lake Tanganyika is crucial for aquatic ecosystems’ productivity, particularly fish resources and human health. It is crucial to preserve high water quality in the Lake Tanganyika region since the lake is a major source of drinking water and irrigation for the local communities.

Moreover, sustainable management practices are crucial to mitigate water quality deterioration and safeguard public health and the ecosystems of Lake Tanganyika. The research is particularly significant as Lake Tanganyika is Bujumbura’s primary water source and is highly polluted [33]. In addition, water scarcity, population growth, urbanization expansion, industrial development, lifestyle changes, and inefficient distribution are major issues in Bujumbura, resulting in sporadic shortages and inconsistent supplies [34]. Accurate prediction of water use can help with resource management and provide a more steady supply of water for the region’s expanding population. Monitoring water quality and consumption is, therefore, essential to lessening the effects of contaminated and filthy water.

Nonetheless, recent studies have effectively used recurrent neural networks and machine learning techniques to predict water quality and consumption and classify water quality forecasting. Depending on this success, the current study leverages the water quality index (WQI) and consumption, incorporating diverse metrics to enhance prediction models. The main contributions of this paper include uniquely employing multi-model architectures such as GRU, LSTM, and BiLSTM, differing from previous single-model approaches. These models capture complex temporal dependencies, improve prediction accuracy and robustness, and offer insights into Lake Tanganyika’s water quality and consumption dynamics. The use of multiple models facilitates thorough comparisons, advancing research in water quality assessment and demand planning.

Additionally, SVM, KNN, and RF are employed for water quality categorization, broadening the study’s methodological scope and providing robust, interpretable results for environmental management. Accurate predictions aid in effective water resource management and pollution control. This pioneering study, focusing on the Lake Tanganyika region, offers valuable contributions to water quality management for Burundi, the Democratic Republic of Congo, Zambia, and Tanzania, ensuring ecological integrity and sustainable resource use.
2. Materials and Methods

2.1. Description of the Study Area

Lake Tanganyika, the largest of the three lakes in East Africa’s Great Rift Valley, spans from 3° S to 9° S and stretches 676 km from north to south, with an average width of 50 km. Situated mainly in the DRC’s southern basin, it boasts a maximum depth of 1471 m and an average depth of 570 m, with a shoreline spanning 1828 km. Covering 32,900 km², it holds 18,880 km³ of water, making it the continent’s largest reservoir, containing 1/6 of the world’s surface freshwater reserve. Its equatorial location fosters a tropical environment with high temperatures and heavy precipitation [35]. Daily temperatures are typically warm to hot, ranging from around 25 °C to 30 °C, fostering a rich biodiversity in aquatic and terrestrial ecosystems. Heavy and frequent daily precipitation over 100 mm is common during the rainy season in the Lake Tanganyika region. While these periods of heavy precipitation support the surrounding ecosystems and help restore the lake’s water levels, they can also cause erosion and flooding in certain locations.

Lake Tanganyika’s unique ecosystem is influenced by various factors, including seasonal precipitation patterns, river influxes, and interactions with the East African Rift System [36]. While nourished by numerous rivers, its drainage area is limited by steep surrounding mountains, with the Malagarasi River being its primary inflow from Tanzania’s east [37] and the Lukuga River serving as its significant outflow to the Congo River watershed. The region experiences distinct wet and dry seasons, with the basin supporting over 10 million people at a density of approximately 43 persons/km², primarily reliant on agriculture. Lake Tanganyika is shared by bordering countries at 12% (Burundi), 43% (DRC), 36% (Tanzania), and 9% (Zambia). The climate of this area indicates that the wet season lasts for the rest of the year, and the dry season lasts from June to September or October. However, the area faces environmental challenges, including pollution from human activities such as deforestation, agricultural runoff, and industrial waste, threatening the delicate balance of this unique ecosystem. In addition, urbanization expansion around the lake and population growth raise concerns about pollution, as all residential waste is currently deposited untreated into the lake. Figure 1 shows the study area and its surrounding waters.

Figure 1. An outline of Lake Tanganyika and its surrounding waters.
2.2. Methodology

This study utilizes artificial intelligence techniques for water quality and consumption forecasting, employing cutting-edge methods. Deep learning models predict water quality index (WQI) and urban water consumption (UWC), while machine learning models like SVM, K-NN, and RF forecast Lake Tanganyika’s water quality categorization. The study offers a straightforward yet efficient forecasting approach, as depicted in Figure 2. Model effectiveness is evaluated using precision, F1-score, accuracy, and recall. $R^2$, Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Nash–Sutcliffe (NSE) also gauge deep learning model performance. Developing highly effective models for water contamination forecasting facilitates prompt solution implementation. The Pearson correlation matrix coefficients (PCMC) have been implemented to evaluate the relationship between water properties. The study outlines data collection, DL and ML model development, and result evaluation in subsequent sections.

![Figure 2. Framework of the adopted methodology in this current study.](image)

2.2.1. Datasets

The dataset for the present study was gathered from the water and electricity production and distribution authority (REGIDESO) historical archives. Water quality measures were the subject of daily data collection from 2018 to 2023. Electrical conductivity (EC), pH, dissolved oxygen (DO), nitrate ($NO_3^-$), phosphates, suspended solids (SS), and water temperature (WT) are the six important metrics that are included in the dataset. The databases also include the appropriate date and the urban water consumption (UWC) records for the city of Bujumbura.

2.2.2. Data Preprocessing

The data processing stage of data analysis is essential to enhancing the data’s quality. At this point, the WQI was determined and normalized using the dataset’s most crucial metrics. Subsequently, the water samples were classified using the WQI values. The UWC data has undergone processing and standardization. The z-score method has been used as a data standardization tool to increase accuracy.

2.2.3. Water Quality Index Calculation

The WQI is determined by considering many parameters that significantly impact water quality. This study utilizes a dataset of six important water quality characteristics
to evaluate the suggested model. The method used is crucial for calculating the weighted arithmetic water quality index (WAWQI), a method initially proposed by Horton in 1965 [38] and further developed by Brown et al. [39]. The provided representation is the weighted arithmetic water quality index (WQI) calculation.

\[
WQI = \frac{\sum_{i=0}^{n} W_i q_i}{\sum_{i=0}^{n} W_i}
\]  

where \(n\) is the number of parameters used in the water computation, \(q_i\) is the quality rating scale of each parameter \(i\) calculated by Equation (2) mentioned below, and \(W_i\) is the weight unit calculated for each parameter, as highlighted in Equation (3) below.

\[
q_i = \frac{(V_i - V_{\text{ideal}})}{(S_i - V_{\text{ideal}})} \times 100
\]  

where: \(V_i\) represents the quantified value of parameter \(i\) in the examined water samples. According to Table 1, \(S_i\) is the recommended standard value for the parameter \(i\). The ideal state is characterized by clean water, with all parameters set to 0 except for pH, which is set to 7.0, and dissolved oxygen (DO), which is set to 14.6 mg/L.

\[
W_i = \frac{K}{S_i}
\]  

Assuming the following calculation for \(K\), this stands for the proportionality constant. It can be calculated as follows:

\[
K = \frac{1}{\sum_{i=1}^{n} S_i}
\]

Table 1. Acceptable ranges for the variables used to determine WQI [40].

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Allowable Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>pH</td>
<td>8.5</td>
</tr>
<tr>
<td>DO, mg/L</td>
<td>10</td>
</tr>
<tr>
<td>(NO_3^-), mg/L</td>
<td>45</td>
</tr>
<tr>
<td>Phosphate</td>
<td>0.1</td>
</tr>
<tr>
<td>SS, mg/L</td>
<td>100</td>
</tr>
<tr>
<td>EC, (\mu S/cm)</td>
<td>1000</td>
</tr>
</tbody>
</table>

Table 2. Water quality classification based on WQI [41].

<table>
<thead>
<tr>
<th>Range of WQI</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–25</td>
<td>Outstanding</td>
</tr>
<tr>
<td>26–50</td>
<td>Good</td>
</tr>
<tr>
<td>51–75</td>
<td>Poor</td>
</tr>
<tr>
<td>76–100</td>
<td>Very poor</td>
</tr>
<tr>
<td>&gt;100</td>
<td>Not suitable to drink</td>
</tr>
</tbody>
</table>

Table 3. Weights according to parameters used in this current study.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Weight Unit ((W_i))</th>
</tr>
</thead>
<tbody>
<tr>
<td>pH</td>
<td>0.011476789</td>
</tr>
<tr>
<td>DO</td>
<td>0.00975527</td>
</tr>
<tr>
<td>Nitrate</td>
<td>0.002167838</td>
</tr>
<tr>
<td>Phosphate</td>
<td>0.975527024</td>
</tr>
<tr>
<td>SS</td>
<td>0.000975527</td>
</tr>
<tr>
<td>EC</td>
<td>9.75527 \times 10^{-5}</td>
</tr>
</tbody>
</table>

\[
w_i = \frac{K}{S_i}
\]

Assuming the following calculation for \(K\), this stands for the proportionality constant. It can be calculated as follows:

\[
K = \frac{1}{\sum_{i=1}^{n} S_i}
\]
2.2.4. 1 Z-Score Normalization Process

Simplifying computations can be achieved through normalization. It turns from a dimensional expression into a scalar and a no-dimensional expression. When normalizing parameters (also called normalization scores) are applied, the Z-score normalization method uses the $\mu$ and $\sigma$ values of the data tested. It is important to note that the $z$-score is important for standardizing the dataset because it normalizes data by centring it around the mean with a standard deviation of one, ensuring that all features contribute equally to the analysis and improving the performance of machine learning algorithms sensitive to feature scales. The following formula is implemented to compute it:

$$Z - \text{score} = \frac{(x - \mu)}{\sigma}$$

where: This represents an individual data point

$\mu$ (mu): This symbolizes the mean, or average, of the dataset.

$\sigma$ (sigma): This denotes the dataset’s standard deviation.

### Water Quality Index Calculated

The results and trends discovered following Lake Tanganyika’s daily WQI calculation from 2018 to 2024 are displayed in Figure 3. The WQI has been computed using WAWQI techniques [42]. The acceptable criteria and weights for the water quality index have been established and strictly adhered to. The continual discharge of contaminants into water bodies from industrial sources, agricultural runoff, heavy rainfall, and inadequate sewage treatment systems are most likely the main reasons for the daily increase in the water quality index for 2023–2024 [43]. Overall water quality gradually decreases over time due to this ongoing pollution, which raises contamination levels [44]. Growing numbers of people living in the bordering cities of Lake Tanganyika, intensifying agriculture, and expanding industries will all lead to increased pollution and declining water quality [45]. The increased water quality index (WQI) in Lake Tanganyika is attributed to factors such as rising temperatures due to climate change, leading to thermal stratification, municipal wastewater, stormwater, improper waste disposal, oil spills and chemical spills, construction waste, atmospheric deposition, and decreased oxygen levels [46]. Additionally, increased anthropogenic activities, including industrial pollution, contribute to nutrient loading and eutrophication, further degrading water quality. An increase in the water quality index (WQI) typically indicates deteriorating water quality, which can complicate predictions by introducing more variability and uncertainty into water quality models. It necessitates more sophisticated and adaptive predictive techniques to accurately capture the effects of pollution, climate change, and human activities on water resources. Proactive environmental laws, financial support for wastewater treatment facilities, and neighborhood-based conservation initiatives could reduce pollution and raise the WQI daily (Figure 3).
2.2.5. Urban Water Consumption

Urban water consumption forecasting uses a dataset of UWC and dates. The consumption variable holds the water utilized on that date, whereas the “Date” column has a timestamp. It is worth noting that the dataset includes data from 2018 through 2024. Figure 4 demonstrates the slow variation in water demand in Bujumbura from 2018 to 2020. This variation is observed due to the natural replenishment of water sources, reduced extraction needs, and decreased outdoor water usage. Figure 4 displays the demand in Bujumbura using data from the dataset on water consumption predictions. From 2020 to 2023, urban water consumption is rising due to population expansion, urban area development, climate change, and lifestyle changes that drive the need for water-intensive activities like landscaping and household use [47]. From 2023 to 2024, the highest increase in urban water consumption in Bujumbura city is primarily driven by rapid population growth, which escalates domestic, agricultural, and industrial water demand.

![Figure 4. Urban water consumption of Bujumbura City from 2018 to 2024.](image)

Additionally, urbanization and economic development have intensified the need for improved infrastructure and services, further amplifying water usage [48]. However, the increase in urban water consumption in Bujumbura city strains existing water resources, leading to potential shortages and heightened competition for water among the agricultural, industrial, and residential sectors. This case can exacerbate predictions of water scarcity, potentially accelerating the timeline for critical shortages and necessitating more urgent water management and conservation measures.

On the other hand, water-saving measures, seasonal variations in the weather, and the implementation of water-efficient practices and technology have also impacted daily variations in urban water use. Nevertheless, water demand has varied due to water conservation awareness and water-saving technology improvements from 2020 to 2023. In addition, the gradual decrease in consumption may continue to be observed when the measures of water conservation awareness and water-saving technology improvements implemented continue in the next few decades (Figure 4). Water demand may continue to rise due to growing urbanization, industrialization, and population growth, necessitating the strongest management measures.
2.2.6. Machine Learning and Deep Learning

The basic theory of ML, like the support vector machine, K-nearest neighbors, and random forest model, has been implemented. The DL, like Gated Recurrent Unit (GRU), Long Short-Term Memory (LSTM), and Bidirectional Long Short-Term Memory (BiLSTM) neural networks, are also covered in this section. It also examined whether creating a WQI and UWC time series data prediction system is feasible and has a sound theoretical basis.

Support Vector Machine

The SVM was established by Corinna Cortes and Vladimir Vapnik in 1995 [49]. SVMs have demonstrated efficacy across several areas because they can identify ideal hyperplanes for class separation in feature space, even in datasets with high dimensions. The contributions of Cortes and Vapnik have had a profound influence on the domain of machine learning, facilitating progress in pattern recognition and data analysis. It can be enlarged to fit different machine learning issue models. The requisite coefficients can be derived by splitting the points of the input vectors using the hyperplane. Support vectors are the points within the hyperplane. Equation (6) utilizes the Gaussian radial basis function and the linear SVM model to categorize the assessed water samples according to their quality.

\[ K(X, X') = \exp\left(-\frac{\|X - X'\|^2}{2\sigma^2}\right) \]  

(6)

The 2 feature inputs squared Euclidean distance (SED) is denoted by the \(\|X - X'\|^2\) vectors of the feature of the input set, which are represented by \((X, X')\). A free parameter is denoted by \(\sigma\).

K-Nearest Neighbors Model

The KNN is a key ML technique for regression and categorization tasks. When using the KNN to forecast data, the algorithm selects the closest K \((P)\) \(\) neighbors \(\) from the training dataset using a specified metric, such as Euclidean distance. Tested data can be categorized since most results fit into one group. Given that it is utilized to get the neighbors \((p)\) in the vectors of features closest to one another, the value of K would be distinct. The formula below can express \((Di)\).

\[ D_i = \sqrt{(x_1 - x_2) + (y_1 - y_2)^2} \]  

(7)

where: \(x_1, x_2, y_1, \) and \(y_2\) represent the variables in the input data.

Random Forest model

The RF machine learning model is based on trees and combines the results of several DT fits using training data that is selected at random [50]. For every DT in RF, an indication such as the gini-index or information gain is used to create the root node. For both classification and regression issues [51]. The RF technique uses two randomization sources. The “root node” or “parent node” selects the dividing input characteristic at random to produce the first source [52]. Bootstrapping is another widely used method that introduces unpredictability into the RF-generating process. It is explained as randomly dividing the training data into tiny subsamples and then utilizing them to build multiple trees. Using bootstrap sampling techniques, RF solves the overfitting issue [53]. RF was deployed using two hyperparameters. Three hundred DTs were used in the prediction procedure when the n-estimator variable was assigned a value of 300. The 50 set as the max-depth option indicates the greatest depth to which a tree can grow. The RF approach involves combining a set of “base learners”, \(h_1(x)\ldots h_j(x)\), to form the “ensemble predictor”, \(f(x)\) [54]. Equation (8) states that the average of each tree’s prediction in the model yields the outcome for the regression issues.
Although there are several criteria for calculating the model result error, the mean squared error is typically employed. The model is trained until the mean squared error reaches its lowest value. A majority vote is used to decide the forecast for the categorization tasks. The final estimated output is displayed in Equation (9).

\[ f(x) = \frac{1}{J} \arg \max \sum_{j=1}^{J} I(y = h_j(x)) \]  

The RF model is trained until it reaches the lowest value of the splitting criterion, which could be entropy, the Gini index, or the misclassification error. Equation (10) states that the Gini index is the splitting criterion utilized in the current model. The split input feature in the “root node” is selected randomly. The feature with the lowest Gini index is the next to be divided, and it computes the impurity in the “child node” associated with each input feature [55]. The Gini index is zero after each node is split into two. “Terminal nodes” have a zero Gini value [56].

\[ \text{Gini Index} = \sum_{k \neq k}^{K} P_k P_k; \quad \hat{P}_k = \frac{1}{n} \sum_{i=1}^{n} I(y_i = k) \]  

The RF model is trained until it reaches the lowest value of the splitting criterion, which could be entropy, the Gini index, or the misclassification error. Equation (10) states that the Gini index is the splitting criterion utilized in the current model. The split input feature in the “root node” is selected randomly. The feature with the lowest Gini index is the next to be divided, and it computes the impurity in the “child node” associated with each input feature [55]. The Gini index is zero after each node is split into two. “Terminal nodes” have a zero Gini value [56].

Long Short-Term Memory (LSTM Model)

LSTM networks, similar to other recurrent neural networks (RNNs), usually sequentially handle time series data, processing one step at a time. Consequently, the model generally lacks access to future data points when making predictions or learning a sequence of data points [58]. In the context of time series forecasting or sequence prediction tasks, this attribute is perceived as a constraint since the model lacks immediate access to forthcoming data that could enhance the accuracy of predictions. Consequently, LSTM networks may encounter difficulties capturing specific patterns or relationships that span extended periods, particularly if those patterns rely heavily on future data. Nevertheless, despite this constraint, LSTM networks can capture and acquire intricate temporal patterns from past data. Their capacity to retain a memory state over some time enables them to capture relationships within the sequence up to the present moment. In addition, although not directly, instructor pushing and attention processes can assist LSTM networks in indirectly incorporating information from future time steps. Although LSTM networks do not explicitly incorporate future data in their processing, they can still be highly effective for various tasks such as time series forecasting and sequence prediction [59]. This is especially true when trained on extensive datasets containing ample past information. Precursory RNNs were heavily utilized to process series data.

However, due to the recurring gradient vanishing and explosion issues, RNN is no longer appropriate for making long-term predictions in time series data. Long short-term memory networks (LSTM) and RNN networks with gates and functions evolved, and the concept of cell state was first implemented. LSTM networks are an improved form of RNN networks that are well-suited for managing long-time series data. When it comes to time series data prediction, one of the best RNN algorithms is the LSTM model. The LSTM model uses a logistic sigmoid as its activating function. Provided the input gate is closed and the forget gate is open, the memory cell keeps reminding the user of their initial entry, resolving the common RNN issues [60]. The formulas for the RNN model are as follows:

\[ h_t = \tanh (w_i \cdot h_t + w_x x_t) \]
\[ y_t = \omega_y \cdot \omega_t \]  
\hspace{1cm} (12)

where \( h_t \) denotes the neural network’s hidden layer for the given training set of data \((x_t)\). \( y_t \) is a representation of the output layer. Nevertheless, \( \omega_t \) and \( \omega_y \) are the brain cells and matrix’s weights. The three main parts of an LSTM are the input gate, the forget gate, and the output gate. The LSTM model is computed using the following formulas:

**Input gate:**
\[ i_t = \sigma \left( W^{(ii)} \bar{x}_t + W^{(hi)} h_{t-1} \right) \]  
\hspace{1cm} (13)

**Forget gate:**
\[ f_t = \sigma \left( W^{(if)} \bar{x}_t + W^{(hf)} h_{t-1} \right) \]  
\hspace{1cm} (14)

**Output gate:**
\[ o_t = \sigma \left( W^{(io)} \bar{x}_t + W^{(ho)} h_{t-1} \right) \]  
\hspace{1cm} (15)

**Process input:**
\[ \tilde{C}_t = \tanh \left( W^{(iC)} \bar{x}_t + W^{(hC)} h_{t-1} \right) \]  
\hspace{1cm} (16)

**Cell update:**
\[ C_t = C_{t-1} + i_t \tilde{C}_t \]  
\hspace{1cm} (17)

**Output:**
\[ y_t = o_t \cdot \tanh(C_t) \]  
\hspace{1cm} (18)

Python was utilized to examine the LSTM. The LSTM layer contains 23 open variables. To complete the task, decide on the units, invoke the function, provide the sequence back, and terminate the procedure. The structure of LSTM is demonstrated in Figure 5 below:

![Figure 5. Model’s architecture for the LSTM.](image)

Gated Recurrent Unit (GRU)

LSTM and GRU architectures are specifically developed to mitigate the vanishing gradient issue in conventional RNNs [27]. This problem hampers the network’s capacity to capture long-term relationships in sequential input effectively. They accomplish this by implementing systems that allow for the selective retention or forgetting of information over some time [61]. LSTM achieves this by employing a more intricate architecture incorporating three gates: the forget gate, the input gate, and the output gate. These mentioned gates regulate information flow inside the cell, figuring out when changes to the memory are necessary.

In contrast, GRU streamlines the structure by merging the gate’s forget and input into a unified “update gate”. In addition, compared to LSTM, GRU uses fewer parameters by combining the cell state and hidden state. Nevertheless, a GRU has two rather than an LSTM’s three layers [62]. Regarding computing power, GRU is more efficient
than LSTM and vanilla RNN. It is primarily referred to as the conventionalized LSTM model. Figure 6 depicts the GRU structure.

$$\text{Reset gate: } r_t = \sigma(W^{(ir)}\tilde{x}_t + W^{(hr)}h_{t-1})$$  \hspace{1cm} (19)

$$\text{Update gate: } z_t = \sigma(W^{(iz)}\tilde{x}_t + W^{(hz)}h_{t-1})$$  \hspace{1cm} (20)

$$\text{Process input: } \tilde{h}_t = \tanh(W^{(ih)}\tilde{x}_t + W^{(hh)}h_{t-1})$$  \hspace{1cm} (21)

$$\text{Hidden state update: } h_t = (1 - z_t)h_{t-1} + z_t\tilde{h}_t$$  \hspace{1cm} (22)

Output: $$y_t = h_t$$  \hspace{1cm} (23)

![Figure 6. Model’s architecture for the GRU.](image)

**Bidirectional Long Short-Term Memory**

LSTMs are specifically engineered to capture interdependencies within sequences by selectively keeping or discarding information over some time. However, they usually handle sequences linearly, utilizing past knowledge to anticipate future information but not immediately integrating future information during processing. BiLSTM, in contrast, overcomes this constraint by analyzing the input sequence in both the forward and backward directions [63]. This feature enables the model to comprehend and depict the full sequence more effectively by incorporating dependencies from both preceding and subsequent contexts. It employs two hidden layers and an output layer that joins them. It stores past and future information on the present-time basis of the time series data. They are stacking forward and backwards LSTM, and a rather robust improved network known as BiLSTM is produced. BiLSTM networks have the potential to foresee a specific time and rely less on pre-series than LSTM networks since they synthesize the input data output from the upper and lower time points. The BiLSTM hidden layer output includes forward and backward activation outputs. The BiLSTM formulation is shown in Equations (19)–(21), where the input for the hidden layer is represented by $H_t$ ($\overline{h}_t$), the activation function is indicated by $\sigma$, and the output is generated by updating the forward structure ($h_t$), and backward structure ($\overline{h}_t$). Figure 7 depicts the BiLSTM structure.

$$\overline{h}_t = \sigma(W_{\overline{h}x}\overline{x}_t + W_{\overline{h}h}\overline{h}_{t-1} + b_{\overline{h}})$$  \hspace{1cm} (24)

$$h_t = \sigma(W_{hx}x_t + W_{hh}h_{t-1} + b_h)$$  \hspace{1cm} (25)
\[ H_t = W_{xh} \vec{h} + W_{yh} \vec{h} + b_y \]  

2.2.7. Evaluation of Performance

There are various error measures available to examine a classification model’s performance. The accuracy score, precision, recall, and F-score are defined by Equations (27)–(30). These metrics are obtained after normalizing the confusion matrix by comparing the predictions with the actual classes. There are five classes in this research, which is noteworthy. A normalized confusion matrix illustrating five classes that account for the outcomes is displayed in Figure 8. The findings after classification are shown in the following table and figures.

\[
\text{Accuracy} = \frac{\text{TN} + \text{TP}}{\text{TN} + \text{FN} + \text{TP} + \text{TN}} \tag{27}
\]

\[
\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{28}
\]

\[
\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{29}
\]

\[
\text{F1 - score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{30}
\]

where: FP, FN, TP, and TN stand for false positive, false negative, true positive, and true negative, respectively.
Figure 8. Normalized confusion matrices for machine learning models showing correct and misclassifications: (a) Support Vector Machine; (b) k-nearest neighbors; (c) Random Forest.

Deep machine learning has been used to predict WQI and UWC, and the following equations are currently used: It is worth noting that MAE, RMSE, $R^2$, and NSE are utilized to assess the robustness of the created models in the present research.

$$\text{MAE} = \frac{\sum \left(\left\| x_o - x_p \right\| \right)}{n}$$  \hspace{1cm} (31)

$$\text{RMSE} = \sqrt{\frac{\sum (x_o - x_p)^2}{n}}$$  \hspace{1cm} (32)
2.2.8. Utilizing the Correlation Matrix

The Pearson’s correlation matrix coefficient (PCMC) method analyzes the correlation between the significant dataset characteristics utilized to forecast the WQI values.

\[
R^2 = 1 - \frac{(\text{Explained Variation})}{(\text{Total variation})}
\]

(33)

where:
- \( x \): represents the input values in the first training dataset
- \( R \): stands for the method of calculating Pearson’s correlation coefficient.
- \( y \): the values entered from the second training set of data;
- \( n \): the total quantity of variables received.

3. Results

This section presents the prediction results from the multi-models of water quality categorization, WQI, and urban water consumption. Though a comparison of the ML currently in use to forecast WQC has been shown via F1-score, precision, recall, and accuracy, a comparative study of DL models used to predict WQI and water demand is presented through \( R^2 \), RMSE, MAE, and NSE. This section has also implemented the Pearson correlation matrix analysis to make it more comprehensive. It is imperative to note that the discussion of the results obtained is implemented in Section 4 below.

3.1. Classification Prediction for the Water Quality of Lake Tanganyika

3.1.1. Normalized Confusion

Since normalized confusion takes into account class imbalances by expressing the matrix in relative terms (percentages or proportions), it is essential for assessing the performance of a classification model. This makes it simpler to compare the model’s accuracy across different classes. Figure 8 of normalization helps to highlight which classes are being predicted accurately and which are not, providing a clearer understanding of the model’s strengths and weaknesses. Figure 8a shows the SVM model’s strong performance, but some errors indicate potential class imbalances or data limitations. Figure 8b shows the KNN classifier excels in some categories but struggles in others. Similarly, Figure 8c shows the Random Forest classifier performs well overall but has trouble distinguishing certain classes.

3.1.2. Performance of Machine Learning for Water Quality Categorization Prediction

The efficacy of Random Forest (RF), K-Nearest Neighbors (KNN), and Support Vector Machine (SVM) machine learning techniques in forecasting water quality classification (WQC) is illustrated in Table 4. This table provides detailed metrics for each method, including precision, recall, and F1-score across different WQC categories, along with overall accuracy, macro average, and weighted average, highlighting that SVM achieved the highest accuracy for testing at 97%, followed by RF at 96%, and KNN at 87%. The table also shows that RF was more effective than SVM and KNN in terms of training accuracy, with a 99.89% success rate.
Table 4. The effectiveness of the machine learning methods utilized to forecast WQC.

<table>
<thead>
<tr>
<th>ML</th>
<th>Metrics</th>
<th>Excellent</th>
<th>Good</th>
<th>Poor</th>
<th>Very Poor</th>
<th>Unifit</th>
<th>Accuracy Test In%</th>
<th>Accuracy Train In %</th>
<th>Macro Avg</th>
<th>Weighted Avg</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Precision</td>
<td>1.00</td>
<td>0.99</td>
<td>0.95</td>
<td>0.85</td>
<td>1.00</td>
<td></td>
<td></td>
<td>0.96</td>
<td>0.96</td>
</tr>
<tr>
<td>RF</td>
<td>Recall</td>
<td>0.75</td>
<td>0.94</td>
<td>1.00</td>
<td>0.75</td>
<td>0.82</td>
<td>96</td>
<td>99.89</td>
<td>0.85</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td>F1-score</td>
<td>0.86</td>
<td>0.97</td>
<td>0.98</td>
<td>0.79</td>
<td>0.90</td>
<td></td>
<td></td>
<td>0.90</td>
<td>0.96</td>
</tr>
<tr>
<td>KNN</td>
<td>Precision</td>
<td>0.67</td>
<td>0.85</td>
<td>0.88</td>
<td>0.85</td>
<td>1.00</td>
<td></td>
<td></td>
<td>0.85</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>Recall</td>
<td>0.50</td>
<td>0.82</td>
<td>0.92</td>
<td>0.73</td>
<td>0.82</td>
<td>87</td>
<td>82.88</td>
<td>0.76</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>F1-score</td>
<td>0.57</td>
<td>0.84</td>
<td>0.90</td>
<td>0.79</td>
<td>0.90</td>
<td></td>
<td></td>
<td>0.80</td>
<td>0.87</td>
</tr>
<tr>
<td>SVM</td>
<td>Precision</td>
<td>0.00</td>
<td>0.96</td>
<td>0.99</td>
<td>0.94</td>
<td>1.00</td>
<td></td>
<td></td>
<td>0.78</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td>Recall</td>
<td>0.00</td>
<td>0.99</td>
<td>0.99</td>
<td>1.00</td>
<td>1.00</td>
<td>97</td>
<td>98.52</td>
<td>0.80</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>F1-score</td>
<td>0.00</td>
<td>0.97</td>
<td>0.99</td>
<td>0.97</td>
<td>1.00</td>
<td></td>
<td></td>
<td>0.79</td>
<td>0.98</td>
</tr>
</tbody>
</table>

3.2. Validation and Settings of the Multi-Model for Water Quality Index Prediction

The number of epochs is critical for training DL like LSTM, BiLSTM, and GRU models as it determines how many times the model iterates through the entire training dataset, allowing for improved parameter tuning. Properly selecting epochs helps balance underfitting and overfitting, ensuring the model generalizes well to new data. Testing the dataset evaluates the model’s performance and prevents overfitting, while training is essential for the model to learn and accurately predict patterns in real-world applications. The findings of the deep learning models with respect to loss per epoch and training and testing daily graphs of the water quality index from 2018 to 2024 are shown in Figure 9. 20% was tested, and 80% of the instruction was completed. For smooth training and validation loss, the learning rate was set at $1 \times 10^{-3}$. Figure 9 illustrates the 50 training epochs that were used to train the multi-model.

![Figure 9](image-url) Deep learning models loss per epoch and training and testing daily graph of water quality index from 2018 to 2024: (a) Bi-directional long short-term memory loss graph per epoch; (b) Long short-term memory loss graph per epoch; (c) Gated recurrent unit and loss graph per epoch; (d) water quality index training and testing daily graph.

The Performance of DL Models for Water Quality Index Prediction

Figure 10 and Table 5 demonstrate the performance of the BiLSTM, LSTM, and GRU models. In Figure 10, the blue line represents the observed daily, and the orange line represents the model’s prediction for 2025 and the first few months of 2026. Forecasted and observed WQI indicate that there’s a chance that summertime water pollution levels will stay high. Table 5 demonstrates that among the deep learning models used to fore-
cast water quality index (WQI), the GRU model achieved the best performance with the lowest MAE (0.3975%), RMSE (0.6941%), and the highest $R^2$ value (0.78).

Figure 10. Performances of models in predicting the daily water quality index of Lake Tanganyika: (a) Bi-directional long short-term memory graph performance in predicting WQI with orange and blue lines; (b) Long short-term memory graph performance in predicting WQI with blue and orange lines; (c) Gated recurrent unit graph of performance in predicting WQI with orange and blue lines.

Table 5. shows the effectiveness of DL models utilized to forecast WQI.

<table>
<thead>
<tr>
<th>Models</th>
<th>MAE (%)</th>
<th>RMSE (%)</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BiLSTM</td>
<td>0.4197</td>
<td>0.7126</td>
<td>0.73</td>
</tr>
<tr>
<td>LSTM</td>
<td>0.4826</td>
<td>0.7624</td>
<td>0.69</td>
</tr>
<tr>
<td>GRU</td>
<td>0.3975</td>
<td>0.6941</td>
<td>0.78</td>
</tr>
</tbody>
</table>
3.3. Validation and Settings of the Multi-Model for Water Demand Forecasting in Bujumbura City

The number of epochs is crucial for training deep learning models like LSTM, BiLSTM, and GRU, as it defines how often the model goes through the entire training dataset, enabling better parameter optimization. Choosing the right number of epochs helps avoid both underfitting and overfitting, ensuring the model performs well on new data. Assessing the model with a test dataset measures its performance and prevents overfitting, while training allows the model to learn and accurately predict patterns in practical applications. Figure 11 illustrates the loss per epoch and the training and testing daily graphs for urban water consumption in Bujumbura from 2018 to 2024, with 80% of the data used for training and 20% for testing. To ensure smooth training and validation loss, the learning rate was set at $1 \times 10^{-3}$. Figure 11 shows the 50 epochs used to train the multi-model.

![Figure 11](image)

**Figure 11.** Deep learning models loss per epoch and testing and training daily graphs of urban water use from 2018 to 2024: (a) Bi-directional long short-term memory loss graph per epoch; (b) Long short-term memory loss graph per epoch; (c) Gated recurrent unit and loss graph per epoch; and (d) water demand training and testing daily graph.

The Performance of DL Models for Urban Water Demand Prediction

Figure 12 and Table 6 highlight the performance of the BiLSTM, LSTM, and GRU models. In Figure 12, the blue line shows the observed daily water demand, while the orange line depicts the model’s predictions for 2025 and the initial months of 2026. Both the forecasted and observed data suggest that water demand during the summer may exceed that of the wetter seasons. Table 6 indicates that the GRU model outperformed the others, achieving the lowest MAE (374), RMSE (530), and highest $R^2$ value (0.81).
Figure 12. Performances of models in predicting daily water consumption in Bujumbura city: (a) Bi-directional long short-term memory graph performance in predicting urban water use with orange and blue lines; (b) Long short-term memory graph performance in predicting urban water demand with blue and orange lines; (c) Gated recurrent unit graph of performance in predicting water demand with orange and blue lines.

Table 6. The performance of DL models to predict urban water consumption.

<table>
<thead>
<tr>
<th>Models</th>
<th>MAE</th>
<th>RMSE</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>BiLSTM</td>
<td>500</td>
<td>726</td>
<td>0.65</td>
</tr>
<tr>
<td>LSTM</td>
<td>461</td>
<td>675</td>
<td>0.70</td>
</tr>
<tr>
<td>GRU</td>
<td>374</td>
<td>530</td>
<td>0.81</td>
</tr>
</tbody>
</table>

3.4. Nash–Sutcliffe Evaluation

The Nash-Sutcliffe Efficiency (NSE) is important for prediction as it measures how well the predicted values match the observed data, with values closer to 1 indicating better predictive performance. Table 7 shows that the GRU model achieved the highest NSE value (0.759) for the water quality index, indicating superior predictive accuracy compared to the BiLSTM (0.699) and LSTM (0.610) models. Table 8 reveals that the GRU model, with an NSE value of 0.720, outperforms both the LSTM (0.650) and BiLSTM (0.590) models in predicting urban water demand.
Table 7. shows the value of NSE for the water quality index.

<table>
<thead>
<tr>
<th>Models</th>
<th>NSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>BiLSTM</td>
<td>0.699</td>
</tr>
<tr>
<td>LSTM</td>
<td>0.610</td>
</tr>
<tr>
<td>GRU</td>
<td>0.759</td>
</tr>
</tbody>
</table>

Table 8. the value of NSE for urban water demand.

<table>
<thead>
<tr>
<th>Models</th>
<th>NSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>BiLSTM</td>
<td>0.590</td>
</tr>
<tr>
<td>LSTM</td>
<td>0.650</td>
</tr>
<tr>
<td>GRU</td>
<td>0.720</td>
</tr>
</tbody>
</table>

3.5. Correlation Matrix

The correlation matrix for water quality parameters such as pH, SS (suspended solids), WT (water temperature), EC (electrical conductivity), DO (dissolved oxygen), nitrate, and phosphates is essential for assessing their collective impact on the water quality index (WQI). It provides insights into how variations in these parameters correlate with changes in WQI, helping to identify key contributors to water quality degradation or improvement and guiding targeted interventions for environmental management. Figure 13 demonstrates the positive and negative correlations between those parameters. It reveals that variations in these parameters correlate with changes in WQI.

![Correlation Matrix](image)

Figure 13. Establishes the association between important dataset properties.

4. Discussion

4.1. Categorization Prediction for the Water Quality of Lake Tanganyika

4.1.1. Analysis of Normalized Confusion

Normalized confusion matrices are pivotal in water classification, especially in multiclass scenarios like this study’s five-class water quality classification. They offer a precise gauge of model prediction efficacy, vital for informed environmental management decisions, where misclassifications can skew resource allocation and water quality assessments. These matrices unveil insights crucial for refining models, enhancing feature selection, and adjusting parameters to boost accuracy and reliability. For instance, examining Figure 8a reveals SVM model performance, with diagonal dominance indicating robust predictions, albeit with some misclassifications, hinting at potential class imbalances or data insufficiencies. Similarly, Figure 8b illustrates KNN classifier proficiency, notably excelling in the “excellent” and “poor” categories but faltering between
“good” and “very poor”. Likewise, Figure 8c showcases Random Forest classifier parallels, performing well overall but facing challenges distinguishing certain classes. Such analyses underscore the necessity of continual model refinement and data enrichment for optimal water quality assessment and management.

4.1.2. Analysis of the Performance of Machine Learning for Water Quality Categorization Prediction

The outcomes of the machine learning methods utilized are displayed in Table 4. Random Forest (RF) performs consistently well across categories. Excellent water quality is identified with high precision (1.00) but slightly lower recall (0.75), resulting in a balanced F1-score (0.86). RF reliably identifies good quality (F1-score: 0.97) and poor quality (F1-score: 0.98), with stable macro and weighted averages at 0.96. However, it struggles more with poor quality (F1-score: 0.79), although overall performance remains consistent. It is worth noting that the testing and training accuracy of RF are 96% and 99.89%, respectively. K-Nearest Neighbors (KNN) exhibits varied performance. It struggles to identify excellent quality (F1-score: 0.57) but performs adequately in other categories, with balanced F1-scores and consistent macro and weighted averages at 0.87. In addition, the testing and training accuracy of KNN are 87% and 82.88%, respectively. Support Vector Machines (SVM) generally performs well. It struggles with excellent quality (F1-score: 0.00) but effectively identifies other categories with balanced F1-scores (0.97 to 1.00) and consistent macro and weighted averages at 0.98. Likewise, the testing and training accuracy are 97% and 98.52%, respectively. The accuracy outcomes demonstrate the performance of RF compared to the SVM and KNN models. The KNN algorithm, however, has performed the poorest. Although RF has a 99.89% accuracy rate, SVM and KNN have accuracy rates of 98.52% and 82.88%, respectively, for training. Nevertheless, the SVM outperformed the RF and KNN for testing accuracy by 97%, but the outcomes of the RF and KNN for testing accuracy were 96% and 87%, respectively. The RF performed exceptionally better than SVM and KNN based on an accuracy of 99.89% obtained during the training.

4.2. Settings and Validation of Multi-Model

Figures 9 and 11 demonstrate the results for deep learning models regarding loss per epoch and training and testing daily graphs of the water quality index and urban water consumption of Bujumbura from 2018 to 2024. The training reached 80%, and the testing was 20%. Deep learning models require a large dataset for a good fit. Figures 9 and 11a–c demonstrate the training refers to one complete pass of the entire training dataset forward and backwards through the BiLSTM, LSTM, and GRU networks, updating the weights to optimize the model’s performance, often repeated multiple times until convergence. It is worth noting that the Adam optimizer, a standard choice for diverse DL tasks, was employed across all three models. Mean Squared Error (MSE) served as the loss function due to its relevance to RMSE and MAE metrics. Batch size’s impact on model development is minimal. The learning rate was fixed at $1 \times 10^{-3}$ for smooth training and validation loss. Early stopping was implemented rather than a fixed iteration count. The study utilized 50 epochs and a batch size of 32. Figures 9d and 11d show the training and testing of the daily dataset of WQI and UWC from 2018 to 2024. Figures 9d and 11d demonstrate that the training of WQI and water consumption reached 80% and 20%, respectively.

4.3. Analysis of the Water Quality Index Trained and Tested

Figure 9d demonstrates the high variation of WQI in the summer, especially in June, July, August, and September. Lake Tanganyika experiences higher pollution levels in the summer months due to increased human activities such as tourism, fishing, industries, agriculture, and construction waste, leading to greater runoff of pollutants into the lake.
The WQI is high from 2023 to 2024 due to the rising temperatures caused by climate change, leading to thermal stratification, municipal wastewater, stormwater, improper waste disposal, oil spills, chemical spills, atmospheric deposition, and decreased oxygen levels [44]. Additionally, the increased anthropogenic activities, including industrial pollution in Burundi, the Republic Democratic of Congo (DRC), Tanzania, and Zambia, contribute to nutrient loading and eutrophication, further degrading water quality [45].

Additionally, warmer temperatures during these months can exacerbate the growth of algae and other aquatic plants, further contributing to pollution levels [64]. However, the WQI decreases in rainy months like October, November, December, January, February, March, April, and May in Lake Tanganyika due to the dilution of pollution and flushing effects caused by higher precipitation levels. Rainfall helps cleanse the lake by reducing pollutant concentrations through increased water flow and sediment transport, improving water quality during these months [65]. An increase in the Water Quality Index (WQI) usually denotes declining water quality, which adds additional unpredictability and uncertainty to water quality models and makes predictions more difficult. However, the models used in this study performed well in prediction according to the R² and NSE results obtained, despite the highest WQI observed at the end of 2024.

4.4. Analysis of Daily Urban Water Demand in Bujumbura City

Figure 11d depicts no high water demand variation in the rainy season, but the highest is observed in the summer, especially in June, July, August, and September. During rainy months, urban water demand experiences slow variation due to the natural replenishment of water sources, reducing the need for extraction and decreasing outdoor water usage as rainfall fulfills irrigation needs, resulting in stable consumption patterns throughout the city. The high variation in urban water demand during the summer months in Bujumbura is primarily due to increased usage for irrigation, outdoor recreation, and higher consumption for cooling purposes. Additionally, population influx from tourism and seasonal migration further strain the water supply. Furthermore, hotter temperatures in this city and reduced precipitation escalate the need for water, intensifying the demand during this period. Population growth, urbanization, industrialization expansion, lifestyle changes, water-efficient techniques, and technological development also cause the highest water demand. The observed water consumption in a water distribution network (WDN) in Bujumbura city is intricately linked to both water demand and water pressure [66]. According to the Global Gradient Algorithm extension to the distributed pressure-driven pipe demand model, water pressure significantly affects how much water is actually delivered to consumers [67]. In scenarios where pressure is low, water may not reach all endpoints adequately, reducing water consumption despite the high demand observed.

Conversely, high pressure can lead to leaks and bursts, causing water loss and inefficient delivery, highlighting the need for balanced pressure management to ensure reliable and efficient water supply in accordance with actual demand [67]. In the same way, Bujumbura city’s growing urban water use is placing a strain on the region’s water supplies, raising the possibility of shortages and intensifying competition among the residential, commercial, and agricultural sectors for water distribution [68]. This situation may make water scarcity forecasts more likely, which could hasten the time until there are serious shortages and call for more immediate water management and conservation efforts [69]. The gradual decrease is observed due to rainfall, water conservation awareness, and technological advancements. In addition, forecasts indicate that by 2030, 28% of the world’s population could reside in cities with a population of at least one million due to the growth in both the number and size of cities [69].
4.5. Comparison of GRU, LSTM, and BiLSTM Models for Water Quality Index and Water Demand

Gated Recurrent Unit (GRU), Long Short-Term Memory (LSTM), and Bidirectional LSTM (BiLSTM) are advanced neural network architectures designed to handle sequential data, with applications in time series prediction, language modeling, and speech recognition. Here, we compare their performance based on three key metrics: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R-squared ($R^2$).

4.5.1. Analysis and Interpretation of Performance Models for Water Quality Index Prediction

The GRU model outperforms the other models with the lowest MAE (0.3975%) and RMSE (0.6941%) values, indicating the smallest prediction errors. Additionally, its highest $R^2$ (0.78) value demonstrates that GRU explains the most variance in the data. The superior performance of GRU could be attributed to its simpler architecture with fewer gates, which makes it both computationally efficient and effective in capturing essential dependencies within the data. However, the LSTM model, known for its capability to handle long-term dependencies with its three gates (input, forget, and output), shows the highest MAE (0.4826%) and RMSE (0.7624%), indicating the largest prediction errors. The lowest $R^2$ value (0.69) implies that it explains the least variance in the data among the three models. This performance may result from overfitting, insufficient tuning, or the specific characteristics of the dataset, which might favor simpler models like GRU.

Moreover, the BiLSTM model, which processes data in both forward and backward directions, delivers intermediate performance. Its MAE (0.4197%) and RMSE (0.7126%) values are better than those of LSTM but not as good as GRU, indicating moderate prediction accuracy. The $R^2$ value (0.73) is also between those of GRU and LSTM. BiLSTM’s bidirectional architecture helps in capturing context from both past and future data, providing an enhanced understanding of the data at the cost of higher computational complexity.

In conclusion, GRU is the most effective model for this dataset, providing the highest accuracy and reliability with the lowest computational complexity, as indicated by the lowest MAE (0.3975%) and RMSE (0.6941%) and the highest $R^2$ (0.78). BiLSTM offers improved context understanding and better performance than LSTM but does not surpass GRU. LSTM, despite its robust architecture for long-term dependencies, performs the least effectively in this scenario, as evidenced by the highest MAE (0.4826%) and RMSE (0.7624%) and the lowest $R^2$ (0.69). Therefore, GRU is recommended for tasks requiring efficient and accurate sequential data prediction, while BiLSTM can be considered when capturing bidirectional context, which is crucial. Based on Figure 10, Figure 10a of BiLSTM, and Figure 10b of LSTM, Figure 10c demonstrates the superior performance of the GRU model according to the trends observed in the blue line and predicted orange line. In Figure 10, the blue line represents the observed daily, and the orange line represents the model’s prediction for 2025 and the first few months of 2026. The observed and forecasted WQI show that water contamination may remain high for the summer months but decrease if the government implements strict water management policies.

4.5.2. Analysis and Interpretation of Performance Models for Water Demand Forecasts

The GRU model demonstrates superior performance with the lowest MAE (374) and RMSE (530), indicating the most accurate predictions and smallest errors. Its highest $R^2$ value (0.81) suggests that the GRU model explains the most variance in the data, making it the most effective among the three models. The simpler architecture of GRU, with fewer gates compared to LSTM, contributes to its computational efficiency and effectiveness in capturing dependencies in the data. Nevertheless, the LSTM model shows moderate performance, with an MAE of 461 and an RMSE of 675. Although its errors are higher than those of GRU, the LSTM still performs better than BiLSTM. The $R^2$ value of
0.70 indicates that LSTM explains a significant portion of the variance in the data, but not as much as GRU. This performance reflects LSTM’s ability to handle long-term dependencies, although it may require more tuning or be less suited to the specific dataset compared to GRU. Yet, the BiLSTM model has the highest MAE (500) and RMSE (726), indicating the least accurate predictions with the largest errors. Its $R^2$ value of 0.65 is the lowest, suggesting it explains the least variance in the data among the three models. While BiLSTM’s bidirectional architecture allows for capturing context from both past and future data, it comes with increased computational complexity and does not outperform GRU or LSTM in this case.

Definitely, the GRU model is the most effective for this dataset, offering the highest accuracy and reliability, as indicated by the lowest MAE (374) and RMSE (530) and the highest $R^2$ (0.81). LSTM performs moderately well, better than BiLSTM, but not as effectively as GRU. BiLSTM, despite its bidirectional approach, shows the least accuracy and explanatory power, with the highest MAE (500) and RMSE (726) and the lowest $R^2$ (0.65). Therefore, GRU is recommended for tasks requiring efficient and accurate sequential data prediction, while BiLSTM may be less suitable for this specific dataset. In comparison to Figure 12a of BiLSTM and Figure 12b of LSTM, Figure 12c depicts the superior performance of the GRU model based on trends observed and predicted. In Figure 12, the blue line represents the observed daily, and the orange line represents the model’s prediction for 2025 and the first few months of 2026. The performance of the multi-model in the research is more necessary than using one model because the findings are compared to get the fit model.

4.6. Comparison of GRU, LSTM, and BiLSTM Models Based on Nash-Sutcliffe Efficiency

GRU (Gated Recurrent Unit), LSTM (Long Short-Term Memory), and BiLSTM (Bidirectional LSTM) are prominent recurrent neural network architectures used for processing sequential data. We evaluate their performance in the prediction of WQI and water demand using Nash-Sutcliffe Efficiency (NSE), a measure of predictive power and accuracy. An NSE value of 1 denotes perfect performance, where estimation error variance approaches zero.

4.6.1. Analysis and Interpretation of the Performance of Multi-Models for Water Quality Index

The GRU model achieves the highest NSE value of 0.759 in predicting WQI, indicating superior predictive accuracy and reliability. The high NSE suggests that GRU closely matches the observed data, effectively capturing the essential patterns and dependencies within the dataset. The simplicity of GRU’s architecture, with fewer gates than LSTM, contributes to its computational efficiency and robustness, making it particularly well-suited for this specific task. Besides, the LSTM model exhibits moderate performance, with an NSE value of 0.610. It indicates that while LSTM is effective at capturing long-term dependencies due to its complex architecture involving three gates (input, forget, and output), it is less accurate than GRU and BiLSTM in this scenario. The lower NSE suggests that LSTM’s predictions deviate more from the observed data, possibly due to overfitting or the need for further tuning to optimize its performance on the given dataset. However, the BiLSTM model performs better than LSTM but not as well as GRU, with an NSE value of 0.699. This scenario indicates that BiLSTM is moderately effective in predictive accuracy, benefiting from its bidirectional architecture that captures context from both past and future data. However, the increased computational complexity does not translate into superior performance over GRU, as indicated by the lower NSE compared to GRU. BiLSTM’s ability to understand bidirectional context is valuable, but it does not outperform the simpler GRU in this case.

In conclusion, the GRU model stands out as the most effective for this dataset, achieving the highest NSE of 0.759, reflecting its superior predictive accuracy and reliability. BiLSTM, with an NSE of 0.699, offers a balanced performance but does not surpass...
GRU. LSTM, with the lowest NSE of 0.610, demonstrates the least accuracy and reliability among the three models. Thus, GRU is recommended for tasks requiring efficient and accurate sequential data prediction, while BiLSTM can be considered when the bidirectional context is important. LSTM may need further tuning or might be less suitable for this specific dataset.

4.6.2. Analysis and Interpretation Performance of Multi-Models for Prediction of Water Demand

The GRU model achieves the highest NSE value of 0.720, indicating the most accurate and reliable predictions among the three models. The high NSE value signifies that the GRU model closely matches the observed data, effectively capturing the key patterns and dependencies within the dataset. The simpler architecture of GRU, with fewer gates than LSTM, enhances its computational efficiency and robustness, making it particularly effective for this task. Besides, the LSTM model shows moderate performance, with an NSE value of 0.650. This indicates that while LSTM is capable of handling long-term dependencies due to its complex architecture involving three gates (input, forget, and output), it is less accurate than GRU. The lower NSE value suggests that LSTM’s predictions deviate more from the observed data compared to GRU, potentially due to overfitting or the need for more tuning to optimize its performance on the given dataset. However, the BiLSTM model has the lowest NSE value of 0.590, indicating the least accurate predictions among the three models. It suggests that while BiLSTM benefits from its bidirectional architecture, capturing context from both past and future data, it does not perform as well as GRU or LSTM in this case. The increased computational complexity of BiLSTM does not translate into better performance, as indicated by the lower NSE value.

The GRU model definitely stands out as the most effective for this dataset, with the highest NSE of 0.720, reflecting its superior predictive accuracy and reliability. LSTM, with an NSE of 0.650, performs moderately well but not as effectively as GRU. BiLSTM, with the lowest NSE of 0.590, demonstrates the least accuracy and reliability among the three models. Therefore, GRU is recommended for tasks requiring efficient and accurate sequential data prediction, while BiLSTM may be less suitable for this specific dataset. LSTM, although capable, may require further tuning or may not be the best fit for this particular scenario. The values depicted in Tables 7 and 8 of GRU nearing 1 signify minimized variance between predicted and actual values, underscoring the superior performance of this model in this study.

4.7. Correlation Matrix Analysis

The Pearson correlation matrix coefficients help establish connections between dataset properties, aiding in selecting optimal settings and clarifying relationships with water quality index (WQI) data. The correlation matrix quantifies linear relationships between physicochemical characteristics and WQI, showing pairwise parameter correlations, direction, and strength. The ideal positive correlation is 1; the ideal negative is −1.

Figure 13 demonstrates that pH moderately correlates positively with suspended solids (SS) at 0.26, suggesting pH increases SS concentration. WT shows linear relationships with other variables (Figure 13). Figure 13 illustrates positive links between pH, SS/EC, and the inverse of nitrate. DO has a somewhat negative association with EC and a somewhat positive association with nitrate/phosphate. Nitrate is somewhat positively associated with DO and negatively associated with phosphate. Phosphate has a moderately favorable relationship with SS/WQI/EC. The SS, pH, and phosphate are favorably correlated. EC weakly positively relates to pH/phosphate/nitrate but negatively to nitrate/phosphate. Water management relies on understanding these interactions via a correlation matrix.
5. Conclusions

Since clean water (CW) is a vital element in health worldwide and is necessary for human health, water quality (WQ) monitoring is required. Environmental protection greatly depends on the modeling and prediction of WQ. Unlike conventional approaches, deep learning and machine-learning-based systems that use features from the water quality index yield reliable findings for predicting water quality, water demand, and the classification of water quality. This paper proposes effective multi-model architectures of recurrent neural networks, including GRU, LSTM, and BiLSTM, that are used to forecast water quality and consumption with excellent accuracy and robustness rather than relying on a single model, aiming to leverage the unique strengths of each model and facilitate thorough comparisons. The suggested method performed exceptionally well, according to experiments using two datasets. Due to their efficiency in managing nonlinear relationships, high-dimensional data, and varied feature interactions, SVM, KNN, and RF are used for water quality classification. It allows for accurate classification, which is crucial for evaluating water quality and guiding decision-making processes. The Random Forest (RF) model demonstrated high effectiveness in forecasting water quality classification (WQC) with a testing accuracy of 96%, achieving a macro average and weighted average of 0.96 across precision, recall, and F1-score metrics. RF’s precision and recall were particularly strong in the “Excellent”, “Good”, and “Poor” categories, though slightly lower for “Very poor”, indicating consistent performance across most classes. The SVM model achieved the highest testing accuracy at 97% and had a macro average of 0.78 and a weighted average of 0.97, with near-perfect scores in most categories except “Excellent”, where precision was 0.00, highlighting an issue in predicting this specific class. The KNN model, with a lower testing accuracy of 87%, showed decent performance in terms of precision and recall, particularly for the “Poor” and “Unfit” categories, but had a relatively lower macro average of 0.76 and weighted average of 0.87, suggesting less consistent performance across all classes compared to RF and SVM. The training accuracy rates of RF are 99.89%; SVM and KNN are 98.52% and 82.88%, respectively. Nevertheless, based on the accuracy obtained of 99.89%, the RF fared better than the SVM and KNN.

Similarly, GRU emerges as the most effective model for this dataset, achieving the highest accuracy and reliability with minimal computational complexity. It is demonstrated by its lowest MAE (0.3975%) and RMSE (0.6941%) and the highest R² (0.78) and NSE of 0.759 for WQ prediction. BiLSTM, which offers enhanced context understanding, performs better than LSTM but does not surpass GRU. Although LSTM is designed for handling long-term dependencies, it performs the least effectively in this context, as shown by the highest MAE (0.4826%), RMSE (0.7624%), and the lowest R² (0.69) and NSE of 0.610. The BiLSTM model, processing data in both forward and backward directions, shows intermediate performance. Its MAE (0.4197%) and RMSE (0.7126%) are better than LSTM but not as good as GRU, reflecting moderate prediction accuracy. The R² value (0.73) and NSE of 0.699 also fall between those of GRU and LSTM. The anticipated and original value of WQI showed relatively little variation, as indicated by the Nash-Sutcliffe efficiency (NSE) score of GRU obtained. In the same way, the GRU model outperforms others with the lowest MAE (374) and RMSE (530) and the highest R² (0.81) and NSE of 0.720 for water demand, indicating the most accurate predictions and smallest errors. Its simpler architecture enhances computational efficiency and effectiveness. The LSTM model, with an MAE of 461 and an RMSE of 675, shows moderate performance and an R² of 0.70 and an NSE of 0.650, reflecting its capability to handle long-term dependencies but being less suitable for this dataset compared to GRU. The BiLSTM model has the highest errors (MAE 500, RMSE 726) and the lowest R² (0.65) and NSE value of 0.590, making it the least effective for this dataset. The GRU model outperforms the LSTM and BiLSTM for both WQI and water demand prediction, according to the R² and NSE obtained. It has been found that, when compared to more intricate deep learning models like BiLSTM, LSTM, and GRU, simpler GRU recurrent neural network designs perform better. The
suggested method performed exceptionally well, according to experiments using two datasets.

The WQI and urban water demand are forecasted using LSTM, GRU, and BiLSTM because they can identify short- and long-term dependencies in time-series data. This makes the forecasts accurate, which is important for managing and controlling WQ, planning water demand, protecting public health, and maintaining ecosystems. This allows for accurate classification, which is crucial for evaluating water quality and guiding decision-making processes. Using multi-model architectures is crucial for maximizing predictive accuracy, effectively capturing diverse data types, and advancing environmental monitoring and management research.

In addition, apart from emphasizing the need for continuous endeavors to oversee and regulate the quality of the local water supply, the article provides valuable insights into the contamination of the water in this vicinity. Compared to the wet season, the dry season has more severe pollution in Lake Tanganyika based on the increased and decreased WQI observed. Urban water demand is also high during the summer months compared to the wet season.

The increased water quality index (WQI) correlated with a significant rise in urban water consumption over the corresponding period in the study. This relationship suggests that as urban water usage increases, so does the potential for pollutants to enter water sources, leading to degradation in water quality. The influx of contaminants from various anthropogenic activities, such as industrial discharge and urban runoff, likely contributed to the deterioration in water quality, as reflected by the elevated WQI values. Therefore, the study indicates a direct association between heightened urban water consumption and deteriorating water quality, highlighting the need for sustainable water management practices to mitigate pollution and preserve water resources. Consequently, particular mitigation actions are required to stop further water quality degradation. These include continuous environmental monitoring, public awareness campaigns, and the establishment of stringent norms for the use and upkeep of lakes. The highest urban water demand is also observed during the dry season, necessitating increased awareness of water conservation and advancements in water-saving technology to progressively lower use. The correlation matrix helps identify important variables causing water quality degradation or improvement. It directs targeted activities for environmental management by providing insights into how variations in these parameters connect with changes in WQI. The negative and positive correlations are observed based on the findings illustrated in Figure 13. One limitation of this research is the variability and limited availability of high-quality, long-term environmental data specific to this region.

Additionally, the complex interactions between diverse ecological, climatic, and human factors make it challenging to create models that accurately capture and predict the lake’s dynamic conditions. Overfitting of the model could ensue. Using the smallest sample size that can be obtained for the trials is one approach, but doing so would leave too few samples for efficient testing and training. As a result, we intend to expand the dataset in our upcoming studies. Further research is needed to forecast and project WQ and urban water demand while taking climate change into account. This will help identify the issues that Lake Tanganyika has faced and will face in the coming decades. A study regarding the water distribution network (WDN) of Bujumbura City is needed to identify the impact of water pressure on its water consumption.

**Author Contributions:** Conceptualization, A.N.; methodology, A.N.; software, A.N. and M.I.; validation, Y.G. and D.Z.; formal analysis, A.N.; investigation, A.N.; data curation, A.N. and P.H.; writing—original draft preparation, A.N.; writing—review and editing, M.I. and A.K.G.; visualization, Z.W. and B.N.; supervision, D.Z. All authors have read and agreed to the published version of the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no funding.
Data Availability Statement: The corresponding author can provide the data necessary to substantiate the conclusions presented in this inquiry upon reasonable request.

Acknowledgments: The authors would like to thank REGIDESO for their assistance in obtaining the urban water consumption and WQ datasets.

Conflicts of Interest: The authors declare no conflicts of interest.

References


Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.