Comparison of Multiple Linear Regression and Artificial Neural Network Models in retrieving Water Quality Parameters using Remotely Sensed Data: Lake Victoria (Tanzanian) Water

Risper Aroko, Dorothea Deus* , Zakaria Ngereja
Department of Geospatial Sciences and Technology, Ardhi University, P.O Box 35176, Dar-es-Salaam, Tanzania
*Corresponding author: dorothea.deus@aru.ac.tz

DOI: https://dx.doi.org/10.4314/sajg.v14i1.11

Abstract:

Water is an essential resource for the survival and well-being of humans and ecosystems; hence, the quality of water is also crucial. To determine the quality of surface water, water quality parameters are traditionally measured by using in-situ measurements. However, accessing such measurements is time-consuming and labor-intensive. Furthermore, it is almost impossible to obtain measurements of the entire waterbody through this method. Therefore, in this study, we compare the capability of Multiple Linear Regression (MLR) and Artificial Neural Network (ANN) models in retrieving water quality parameters from remotely sensed data in inland waterbodies in the case of Lake Victoria, in Tanzania. The models are commonly used for retrieving remote sensing-based water parameters. The performance of MLR and ANN in retrieving Turbidity and Total Dissolved Solids (TDS) data is evaluated. Surface reflectance values from Landsat 8 Operational Land Imager (OLI) sensor images and in-situ data are used to find reliable relationships between Turbidity and Total Dissolved Solids (TDS). The results indicate that the ANN model performs better than MLR in retrieving Turbidity and TDS data: ANN had an accuracy (R^2) of 88.73% and 83.36%, respectively, while MLR had an accuracy (R^2) of 66.66% and 78.42%, respectively. Other criteria that were used for comparison include the standard error (SE), root mean square error (RMSE) and mean absolute error (MAE) which indicated that ANN performed better than MLR. The general distribution of Turbidity and TDS data, as mapped in Lake Victoria, shows that the water quality of the lake, as described by World Health Organization (WHO) standards, is good and could, therefore, be used for human consumption. Based on the results for Turbidity and TDS obtained in this study, we recommend that ANNs and Landsat 8 OLI satellite images be used for water quality parameter modeling.

Key words: *Multiple Linear Regression, Artificial Neural Network, Turbidity, Total Dissolved Solids, Water quality*

1. Introduction

Lakes are vital resources which play an important role in providing various ecosystem services (Bonansea *et al*., 2019). They provide water as an essential resource for the survival and well-being of people and ecosystems such that the quality of water constantly remains crucial(Liu *et al*., 2024a). However, anthropogenic activities have led to the deterioration of water quality in more than half of these systems worldwide (Dai *et al*., 2017; Ji *et al*., 2021; Sidabutar *et al*., 2017; Thai-Hoang *et al*., 2022; Zhu *et al*., 2022). Water quality is a term generally used to describe and define the physical, chemical, and biological characteristics of waterbodies and to identify the possible contamination source that degrades the quality of water (Al-Mukhtar *et al*., 2020; Al-Mukhtar & Al-Yaseen, 2019a; Ritchie *et al*., 2003). Degradation in the quality of water resources emanates mainly from the discharge of waste, pesticides, heavy metals, nutrients, microorganisms, and sediments (Barrett & Frazier, 2016; Thai-Hoang *et al*., 2022). Several water quality standards have been laid down to aid in checking the extent of water pollution, and thus to maintain these water quality standards (EWURA, 2020; URT, 2018; WHO, 2011). Water quality monitoring is crucial for preserving ecosystem health and for supporting the livelihoods of local communities. By monitoring the quality of surface water, it is possible to establish an indicator of the state of health of the surface waterbody in question (Srivastava *et al*., 2020). Such monitoring is also vital for detecting undesirable changes in water quality. Therefore, it is essential to implement best practices and to make concerted efforts to monitor and enhance water quality. In monitoring surface water quality, the distribution of the water quality parameters applicable to the entire waterbody must be known.

Traditionally, field campaigns are conducted to take *in-situ* measurements of water samples at water monitoring stations. The number and location of the sampling stations should be representative of each of the water sources entering the system and representative of the conditions (e.g., dead-ends, loops, storage facilities, and pressure) within each system (Pu *et al*., 2019). However, traditional point sampling methods are not easily able to identify the spatial or temporal variations in water quality, and these methods are time-consuming, labour-intensive, and expensive (Kc *et al*., 2019). Remote sensing is the science and art of measuring an object without being in actual contact with the object (Bhatti, 2008). Remote sensing can be used to acquire water quality parameters in surface waterbodies as snapshots at various temporal resolutions (days, weeks, months, years) (Nakkazi *et al*., 2024; Srivastava *et al*., 2020) Remote sensing-measurement techniques have useful applications in many fields, including the water quality monitoring field (Batina & Krtalić, 2024; Elhag *et al*., 2019; Gholizadeh *et al*., 2016a; Giardino *et al*., 2014; Jin, 2022; Kapalanga, Hoko, & Gumindoga, 2021; Ochaeta *et al*., 2020; Sagan *et al*., 2020). An important principle underlying the use of remotely sensed data is that different objects on the

Earth's surface and in the atmosphere reflect, absorb, transmit, or emit electromagnetic energy in different proportions; it is such differences that allow these components to be identified (Laili *et al.*, 2015).

Monitoring water quality through remote sensing involves first determining a reliable relationship between light reflectance (for certain wavelength bands) and the water parameters collected *in situ* in cases where different relationships between the reflectance and the parameters have developed (Bonansea *et al*., 2019, Tu *et al*., 2018). From the remote sensing perspective, water can be divided into two classes; Case I waters are the ones whereby the optical properties vary according to the phytoplankton concentrations, while Case II waters are those in which the optical properties do not only depend on the phytoplankton concentrations, but also on other constituents such as suspended sediments and dissolved organic matter (Al-Mukhtar *et al*., 2019). With remote sensing, it is possible to have a spatial and temporal view of surface water quality parameters to monitor the waterbodies more effectively and efficiently and to quantify water quality issues (Batina & Krtalić, 2024; Bonansea *et al*., 2019; Gholizadeh *et al*., 2016b; Jin, 2022; Kc *et al*., 2019; Muhoyi *et al*., 2022; Pourghasemi & Gokceoglu, 2019; Ritchie *et al*., 2003; Srivastava *et al*., 2020; H. Yang *et al*., 2022). With rapid environmental change, it is necessary to regularly monitor the quality of water, whereas remote sensing techniques provide more efficient and less labour intensive and cheaper ways to do so (Batina & Krtalić, 2024; Gholizadeh *et al.*, 2016b; Jin, 2022; Kc *et al*., 2019; Muhoyi *et al.*, 2022; Ritchie *et al*., 2003; H. Yang *et al.*, 2022). Remote sensing modeling techniques are usedin the field for extracting various surface water quality parameters (Isık & Akkan, 2024a; Kc *et al*., 2019; Loaiza *et al*., 2023; Palabıyık & Akkan, 2024a; Pu *et al*., 2019). *In-situ* observations are useful for the calibration and validation of remotely sensed estimations of water quality parameters (Gholizadeh *et al*., 2016b). The most commonly applied modeling approaches include multiple linear regression (MLR) and artificial neural network (ANN) modeling (Abyaneh, 2014; Al-Mukhtar *et al*., 2020; Bonansea *et al*., 2019; Isık & Akkan, 2024b; Kwong *et al*., 2022; Liu *et al*., 2024b; Palabıyık & Akkan, 2024a; Pu *et al*., 2019).

However, the performance of these models in Tanzania's inland waterbodies has not been tested. Thus, this topic serves as motivation for this study. Inland waterbodies are important to Tanzania's ecosystems and economy, providing necessary resources such as water, food, and energy. However, there is a significant gap in applying and validating satellite-based water quality models for these unique environments. The existing models, often developed in different regions, may not accurately reflect the local conditions, including varying turbidity levels and seasonal changes to water chemistry and ecological dynamics. This study addresses this gap by adapting and evaluating these models specifically for Tanzanian inland waterbodies. By doing so, it aims to improve local surface water quality assessment and management and offer valuable insights into the use of satellite-based monitoring in diverse and understudied regions. The objective of this study, therefore, is to compare the performance of MLR and ANN in modeling water quality parameters, particularly Turbidity and Total Dissolved Solids (TDS).

1.1. Study Area

Lake Victoria is one of Africa's freshwater Great Lake extending into Tanzania (49%), Kenya (6%), and Uganda (45%), and covering an area of 68,800 km^2 (Nakkazi *et al*., 2024). This study is focused on the Tanzanian water spanning the coastlines of the Kagera, Mwanza, and Mara regions, as indicated in Figure 1. Lake Victoria is the largest lake in Tanzania. It is used for fishing, transportation and for supplying water to the area. The lake has several inlets, with the main inlet being the Kagera river that feeds into the lake on its western margin. The Nile is the only outlet from the lake; it drains into the Mediterranean Sea on the northern coast. The lake lies between longitudes 31°E and 34°E and between latitudes 1°S and 2°S, with the elevation ranging from 920 to 2552 m above mean sea level (Figure 1), and the basin is under the jurisdiction of the Lake Victoria Basin Water Board (LVBWB).

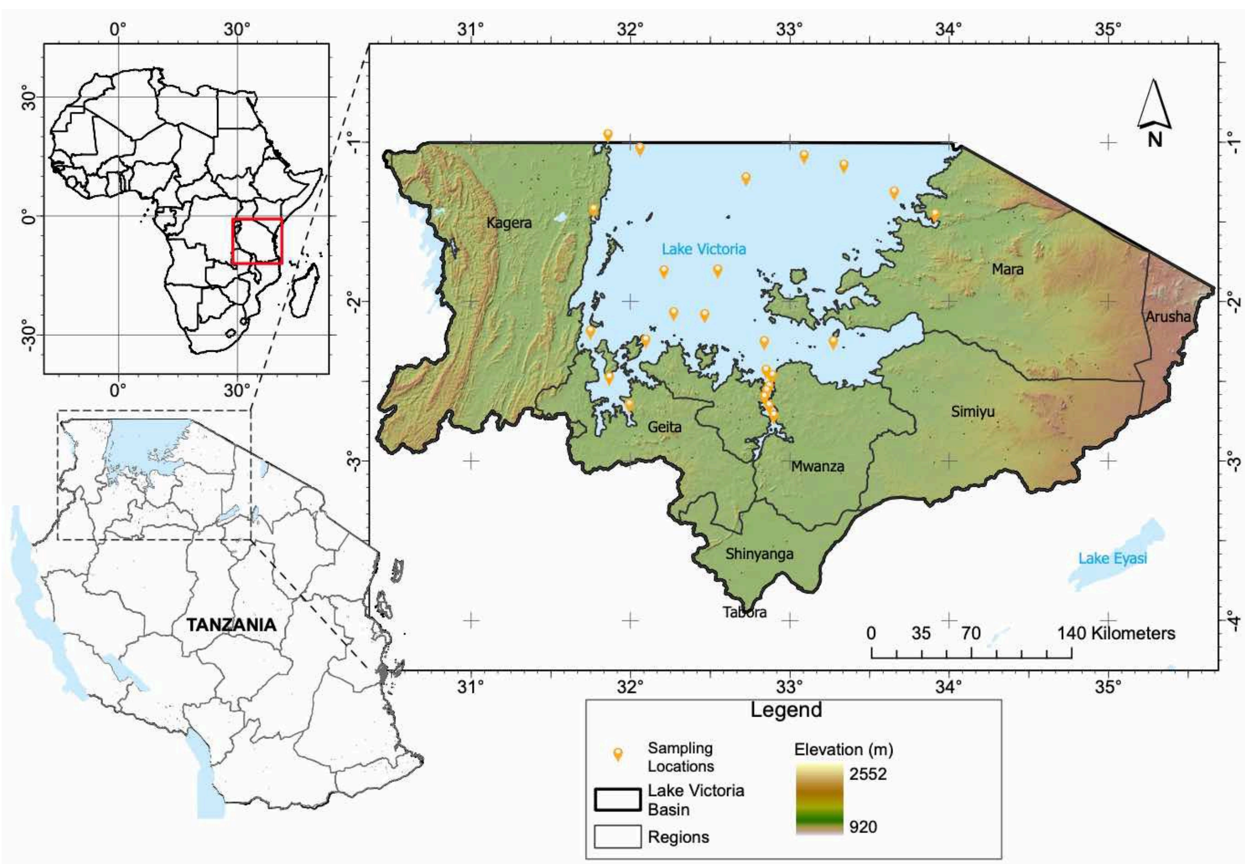

Figure 1: Lake Victoria (Tanzanian water) showing the sampling stations

2. Dataset and Methods

2.1. Data

There are two types of datasets that were used in this study, namely, the remote sensing dataset and the field campaigns dataset. The remote sensing dataset used Landsat 8 OLI images obtained from the United States Geological Survey (USGS) website (https://glovis.usgs.gov/app). The field campaign dataset included the *in-situ* data of Turbidity and Total Dissolved Solids collected from water quality monitoring stations in Lake Victoria's Tanzanian water. About 28 sampling locations were used. All the data used and the supporting dataset are summarized in Table 1. Table 2 presents descriptive statistics of the used water quality data computed for all sampling points (Figure 1), while Figure 2 depicts their distribution before and after normalization using log transformation in the Q-Q plots (Ewuzie *et al*., 2021; Kepford & Warren Hicks, 1993; Mishra *et al*., 2019; G. Yang & Moyer, 2020). The observed water quality data used in this study meet the requirements/standards of the Water Quality Standard of 2016 of the Tanzanian Bureau of Statistics (TBS) , the Energy and Water Utilities Regulatory Authority (EWURA), (URT, 2007), and the World Health Organization (WHO) (EWURA, 2020; URT, 2007; WHO, 2011). According to the WHO Guidelines for drinking water quality, the turbidity level for drinking water should not exceed 1 Nephelometric Turbidity unit (1NTU), and should always be below 1NTU (WHO, 2011). On the other hand, water with a total dissolved solids (TDS) level of less than about 600 mg/l is generally considered to be of a good quality. (Drinking water becomes significantly and increasingly unpalatable at TDS levels greater than about 1000 mg/l.)

Table 1: Datasets and data sources used in the study

Data	Location/ path & row	Acquisition date	Source
Landsat 8 OLI image	170/61	25/06/2020	USGS
	170/62	25/06/2020	(https://glovis.usgs.gov/app)
	171/61	31/05/2020	
	171/62	31/05/2020	
TDS and Turbidity	Lake Victoria	June, 2020	MWAUWASA

Table 2: Observed water quality dataset descriptive statistics (SE =Standard Error, CI = Confidence Interval, TBS -Tanzania Bureau of Standards, TZS =Tanzania, EAS = East Africa, ICS = the document which provides the specifications for water quality sampling)

Descriptive Statistics	Turbidity (NTU)	TDS (mg/l)
Minimum	1.00	43.49
Maximum	10.67	86.36
Range	9.67	42.87
Sum	84.02	1388.22
Median	2.17	51.47
Mean	3.36	55.53
Standard Error	0.58	2.23
Confidence Interval	1.20	4.59
Variance	8.43	123.86
Standard Deviation	2.90	11.13
Coefficient of Variation	0.86	0.20
TBS 2016 STD, TZS 789:2016-EAS 12:2014 ICS 13.060.20	25	1500

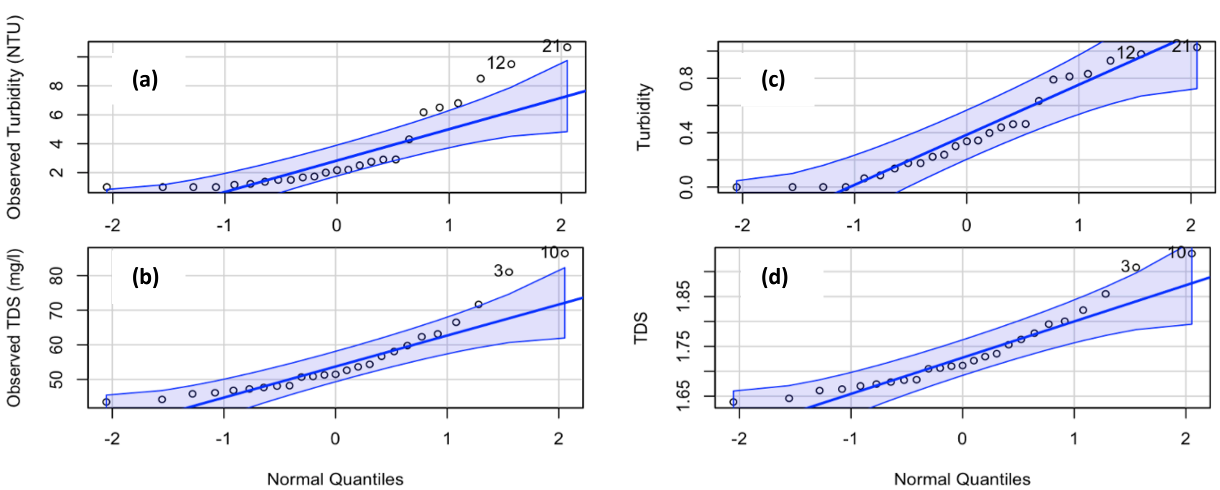

Figure 2: The distribution of the observed (a-b) and log-transformed samples (c-d)

2.2. Methods

2.2.1. Pre-processing of images

The Landsat 8 satellite image used includes four scenes covering Lake Victoria. These images were provided in the form of at-surface radiance levels stored in Digital Number (DN) values. Firstly, atmospheric corrections were carried out to convert the DN values to surface reflectance values using the Semi-automatic Classification Algorithm in Quantum Geographical Information

Systems (QGIS) software. This was carried out individually at all four Landsat 8 scenes by using the Dark Object Subtraction technique, whereby the parameters provided at the metadata file of each image were used (Kapalanga, Hoko, Gumindoga, *et al*., 2021). The atmospheric corrected images were then mosaicked to form a single scene, as indicated in Figure 3. Lake Victoria was then extracted from the Landsat 8 scene by using the boundary of the lake to obtain the surface reflectance of the water area, using only s.

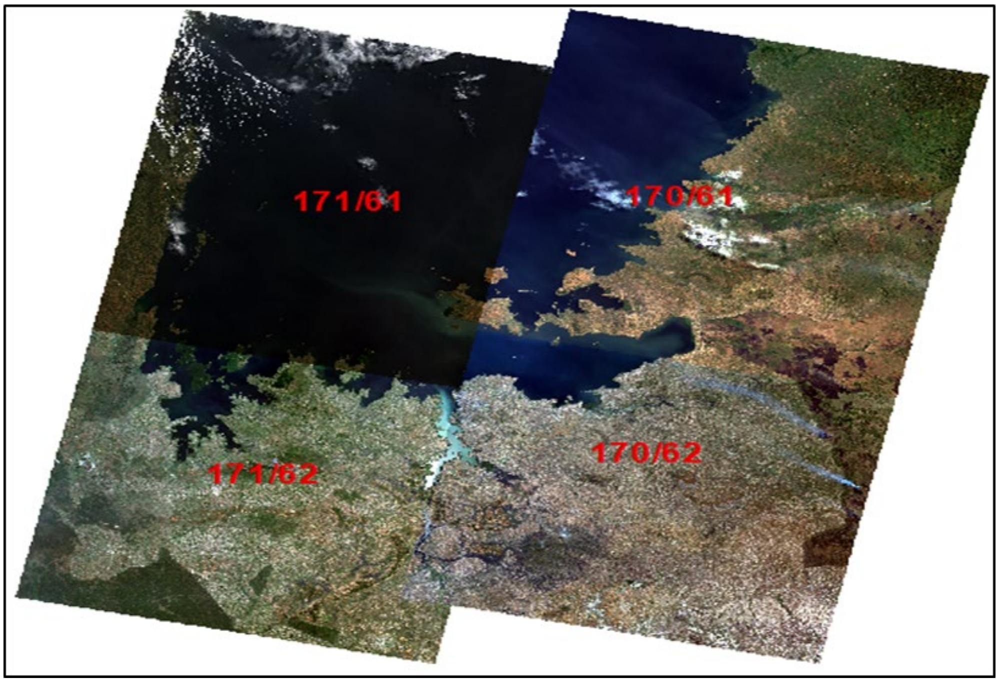

Figure 3: Landsat 8 OLI scenes covering Lake Victoria, as indicated by their paths and rows

Even though the Landsat images are valuable in water quality parameter modeling, they have some limitations in terms of the following categories:

(i) Spatial resolution – Landsat 8 scenes have a spatial resolution of 30 metres, which might be too coarse for smaller or narrower waterbodies, small-scale pollution events, and fine-scaled water quality features (Frasson *et al*., 2024). The 30 m spatial resolution may fail to capture localized variations in water quality, particularly in narrow rivers and small lakes (Huangfu *et al*., 2020).

(ii) Temporal resolution – Landsat has a revisit period of 16 days, which is a relatively long interval and could limit the frequency of observations, especially in sub-Saharan Africa, which is frequently affected by cloud cover. The retrieval of Landsat data may not be frequent enough to monitor rapid changes in water quality.

(iii) Atmospheric interference – Atmospheric conditions such as clouds, haze, and aerosols can affect the quality of Landsat data and lead to challenges in retrieving accurate water quality parameters.

(iv) Depth Penetration: Landsat imagery is limited to surface water observations because the sensor does not penetrate below the surface. Parameters related to subsurface water quality or bottom properties cannot be assessed directly.

(v) Limited Parameters – Landsat data effectively estimate optically active parameters such as chlorophyll-a, turbidity, and suspended solids. However, in the case of non-optically active parameters, such as nutrients, Landsat satellite images are less effective. Generally, it is very difficult to estimate the non-optically active water quality parameters using remote sensing (Arias-Rodriguez *et al*., 2023).

2.2.2. The models used

Multiple linear regression (MLR) is a statistical model used to evaluate the linear relationship between one dependent variable and two or more independent variables (Al-Mukhtar and Al-Yaseen, 2019a). MLR is widely used around the world for estimating water quality parameters (Abyaneh, 2014; Palabıyık & Akkan, 2024b). The general form of the multiple linear regression model is expressed in Equation 1.

$$y = \beta 0 + \beta 1 x1 + \cdots + \beta n * xn + \varepsilon \quad (1)$$

Where y is the dependent variable in the model, $\beta 0$, $\beta 1$, $\beta 2$, $\cdots$ βn are regression coefficients, and x1, x2, $\cdots$ xn are independent variables. ε is the error which follows the normal distribution with $E(\varepsilon) = 0$, and constant variance $Var(\varepsilon) = \sigma 2$.

To perform MLR modeling, the data should be normally distributed and there should not be a multicollinearity problem among the independent variables. As such, the collected *in-situ* data were statistically analysed to check whether the data were normally distributed. The results indicated that the observed data showed a skewed distribution (Figure 2a-b), such that a log transformation was applied to reduce the skewness and normalize them prior to the analysis (Ewuzie *et al*., 2021; Kepford & Warren Hicks, 1993; G. Yang & Moyer, 2020).

A standard practice in remote sensing and environmental modeling is to divide the available data into two subsets: one for calibration (or training) and one for validation (or testing). The calibration subset is used to estimate the model parameters, while the validation subset provides an independent measure of the level of accuracy and robustness of the model. The term 'calibration' in this context refers to the process of adjusting the model parameters to best fit the observed data, thereby ensuring that the model accurately represents the relationship between the satellite observations and the *in-situ* measurements. Thus, a routine procedure for calibrating the model and a representative sample for validating the data were selected. Random splits of 60-80% calibration data and of 20-40% validation data were used (Ali *et al*., 2024; Gholizadeh *et al*., 2016b; Kc *et al*., 2019; Loaiza *et al*., 2023; Mohammed *et al*., 2022; Sagan *et al*., 2020). In this study, about 60% of the *in-situ* Turbidity and TDS data were then used to calibrate the models, with the model then being fitted in the datasets (Mohammed *et al*., 2022). The decision to use 60% of the data for calibration was based on the balance between having enough data to adequately train the model and retaining sufficient data for an independent assessment of the predictive performance of the model. Using 60% of the data for calibration helped ensure that the model could capture the wide range of conditions prevailing in the study area, such as varying levels of turbidity and TDS. This is crucial for developing a robust model capable of accurately predicting these parameters under different environmental conditions. The remaining 40% of the data, which were not used in the calibration process, served as an independent test dataset. This separation helped to avoid overfitting, where the model might have performed well on the calibration data but poorly on new, unseen data.

In model development, the backward elimination method was used. The procedure was initiated by fitting all the possible predictors in the model. The predictors with the highest p-value were then removed and the model was fitted again. This procedure was done repeatedly until all the possible models were generated. Three models for estimating turbidity were generated and three for estimating TDS.

In adopting the best MLR model, the factors that were considered in choosing the model included the coefficient of determination (R^2), the root mean square error of the model, the variation inflation factor (VIF), and the Akaike Information Criterion (AIC). A good MLR model is one that is devoid of a multicollinearity problem between the independent variables (i.e., VIF values 1-5), has the lowest value for AIC, the lowest for RMSE, and a high R^2 (coefficient of determination) (Chan *et al*., 2022; James *et al*., 2021)

Artificial Neural Network (ANN) modeling is a system that simulates the working capabilities of the human brain; it consists of different layers that are connected by weighted links called neurons. This system uses learning phases to model both the linear and non-linear relations of the input and output layers (Isık & Akkan, 2024b; Palabıyık & Akkan, 2024b, 2024a). The architecture of ANN contains a set of approximated mathematical functions which can assume the form (in two ways) of a feed-forward network or a feedback network (W.-B. Chen & Liu, 2015; Y. Chen *et al*., 2020). Essentially, an ANN comprises an input layer, a hidden layer, an output layer, weights, bias, and activation functions (Ritchie *et al*., 2003). The output of an ANN is expressed in Equation 2. Figure 4 depicts the architecture of the ANN model used in this study.

$$\text{output} = \text{sum}\ (\text{weights} * \text{inputs}) + \text{bias} \tag{2}$$

The learning algorithms for the ANN model include gradient descent, the Newton method, conjugate gradient, the Quasi-Newton method, and the Levenberg-Marquardt algorithm (Ciaburro & Venkateswaran, 2017). In this study, a Levenberg-Marquardt feed-forward multilayered ANN was adopted for modelling both Turbidity and TDS, where two input layers were used with three hidden layers having 7, 5 neurons and one (1) neuron, respectively. The initial weights of the input layer were uniform and adaptively distributed with a constant momentum of 0.001. The activation function that was used was Tanh, while the learning epoch was set to 100, and the learning rate annealing of the model was 1×10100 (López *et al*., 2022; Shanmuganathan, 2016). The choice of 100 epochs was based on balancing sufficient training and avoiding overfitting, as the performance of the model stabilized around this point. A learning rate annealing factor of $1{\times}10^{-1001}$ was selected to ensure stable convergence, thereby preventing the model from overshooting the minimum loss function. The momentum was set to 0.001 to accelerate training and smooth out updates, thus preventing oscillations. Other parameters, such as the initial learning rate and batch size, were chosen through experimentation to optimize convergence and computational efficiency.

The training data for the ANN model consisted of *in-situ* measurements of Turbidity and Total Dissolved Solids (TDS) along with corresponding satellite-derived variables. The preparation and processing of these data involved several key steps:

(i) Data collection and preprocessing: *In-situ* data: *In-situ* measurements of turbidity and TDS were collected from various locations on the Lake Victoria waters. These measurements were accompanied by point geolocation data. Satellite Data: Satellite imagery from Landsat was acquired for the same time periods and locations as those for the *in-situ* measurements.

(ii) Data synchronization and alignment: The *in-situ* and satellite data were synchronized to ensure that the satellite observations corresponded to the same timeframes as those for the *in-situ* measurements.

(iii) Feature extraction and selection: The satellite data were processed to extract relevant features, such as surface reflectance values, from specific spectral bands known to be sensitive to water quality parameters. Additional derived indices, such as the Normalized Difference Water Index (NDWI), were also included as potential predictors. Features were selected based on their relevance to turbidity and TDS, and guided by the existing literature and exploratory data analyses.

(iv) Data normalization: All input features and target variables were normalized to a common scale to improve the performance and convergence of the ANN. Normalization was performed using min.-max. scaling, thereby transforming the data to a range of 0 to 1.

(v) Data input format: The prepared data were organized into a tabular format, with rows representing individual observations (sampling points) and columns representing features (satellite-derived variables) and target variables (turbidity and TDS levels). These structured data were then fed into the ANN.

(vi) Data feeding into the ANN: The tabular data were divided into training and test sets, with, as previously mentioned, 70% allocated for training and 30% for testing. The training data were used to optimize the weights and biases of the ANN through backpropagation and gradient descent, thereby minimizing the error between the predicted and actual values. The input layer of the ANN received the satellite-derived features, while the output layer produced predictions for turbidity and TDS. The model architecture was designed to accommodate the number of input features and output predictions, with appropriate hidden layers to capture the complex relationships in the data.

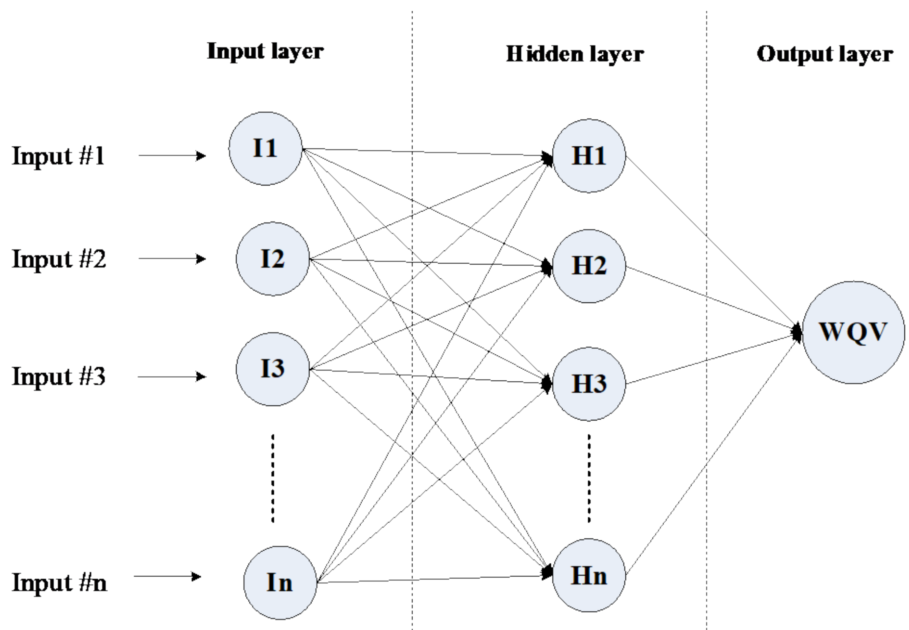

Figure 4: Illustration of the utilized Artificial Neural Network Model (I = input layer, H = hidden layer and WQV = predicted water quality variable)

In order to prevent overfitting in our ANN model, we employed several techniques, and these include:

(i) Train-test Split: We divided our dataset into training and test sets with a 70:30 ratio, thereby ensuring that the model was evaluated on a separate dataset that it had not seen during training.

(ii) Cross-validation: We implemented cross-validation, specifically k-fold cross-validation, where the training data were divided into k subsets. The model was trained k times, each time using a different subset as the validation set and the remaining subsets as the training set. (This method provides a more robust estimate of the model's performance and helps in detecting overfitting.)

(iii) Dropout: Dropout randomly "drops" a fraction of the neurons during training. This prevents the network from becoming overly dependent on specific neurons and forces the model to learn more robust features.

(iv) Early Stopping: Early stopping involves monitoring the model's performance on the validation set during training. If the performance does not improve in a predefined number of epochs, the training is halted to prevent overfitting. This approach helps in finding the optimal number of training epochs.

(v) Data Augmentation: Data augmentation techniques were used to artificially increase the size of the training dataset by applying random transformations (e.g., rotations, scaling) to the training data. This approach helps the model to generalize better by granting it consistency and stability in the face of such transformations.

(vi) Model Complexity: We carefully selected the architecture of the ANN model, including the number of layers and neurons per layer, to balance its complexity and performance. A more complex model can capture more patterns but is also more prone to overfitting. Thus, we chose a model architecture that provided a good trade-off.

(vii) Performance Metrics: We monitored various performance metrics, such as mean squared error (MSE) and R-squared, on both the training and test sets. A significant difference between the performance on these sets could indicate overfitting. By tracking these metrics, we ensured that our model was not just memorizing the training data but also generalizing well.

2.2.3. Comparison of models

To assess the performance of the MLR and ANN models in predicting water quality parameters, the coefficient of determination, R^2, (Tsakiri *et al*., 2018), was determined in terms of Equation 3. R^2 measures the percentage of the variance in the dependent water quality parameter that is explained in terms of the predictor/independent water quality parameters. The coefficient of determination, R^2, ranges from 0 to 1 and is interpreted as a percentage between 0 and 100%; − 0 for poor performance and 100% for strong performance or a sharper level of accuracy in the model. For ideal water quality parameter modeling, the coefficient of determination, R^2, should approach 100% as closely as possible. However, a model performance of R^2= 60% and above is satisfactory.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(O_i - \bar{P}_i)^2}{\sum_{i=1}^{n}(O_i - \bar{O})^2} \tag{3}$$

Where O is the measured/observed water quality parameter and P is the predicted/estimated water quality parameter value. $\bar{O}$ is the mean observed water quality parameter. Other statistical parameters that were used include root mean square error (RMSE) (Equation 4), mean absolute error (Equation 5) and standard error (Equation 6). (Amodel with good performance has a high coefficient of determination, a low root mean square error, a low mean absolute error, and a low standard error (Montgomery *et al*., 2012)).

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_1 - y_i)^2}{n}} \tag{4}$$

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n}|\hat{y}_1 - y_i|}{n} \tag{5}$$

$$\mathrm{SE} = \frac{\mathrm{SD}}{\sqrt{n}} \tag{6}$$

2.2.4. Modeling Turbidity and TDS

The validated models for estimating Turbidity and TDS by using MLR and the ANN model were then used to estimate turbidity and TDS in Lake Victoria to determine their distribution in the study area. For Turbidity, a band combination of optical bands (R*G*B) and a band ratio of green and near infrared (G/NIR) were used. On the other hand, for TDS, a band combination of blue, red, and near infrared bands (B*R*NIR), together with Normalized Difference Water indices (NDWI), was used (Akbar *et al*., 2010). The 70/30 train/test data split for the ANN model was chosen to provide a robust evaluation framework, ensuring that its performance could be assessed on independent data and not just on the data used for training (Sagan *et al*., 2020).

Tanzania's long rainy season, ending in May/June 2020, was marked by increased runoff and sediment influx, which altered the quality of the water, thereby affecting the turbidity levels and TDS of its waterbodies. The elevated turbidity during this period could have interfered with the accuracy of the satellite-based models for water quality mapping. Our study addressed these challenges by timing the data collection and model application phases to account for the impact of the high turbidity levels on the remote sensing signals. Ground truth data validated the accuracy of the models. This highlights the necessity of season-specific calibration and validation to ensure reliable remote sensing-based water quality monitoring in regions such as Tanzania which present with significant seasonal variations.

3. Results

3.1. Developed Models

3.1.1. Multiple Linear Regression

The *in-situ* data used were normalized by applying the log transformation function, as depicted in Figure 2. Shown by the VIF values, 1 – 5, in this figure, the independent variables had no multi-collinearity problem (). The model that was adopted for modeling Turbidity (Equation 6) and TDS (Equation 7) had VIF values of 1 - 5, with the lowest AIC value, the lowest RMSE value and a high R^2 value.

$$\text{Turbidity} = 26485.26 * \text{RGB} + 8.053\frac{\text{G}}{\text{NIR}} - 1.753 \quad (6)$$

$$\text{TDS} = 69841.552 * \text{BRNIR} + 5.127\text{NDWI} + 52.609 \quad (7)$$

The validation of these two models, as indicated in Figure 5, shows that the models have allowable accuracy and can be used for modeling Turbidity and TDS.

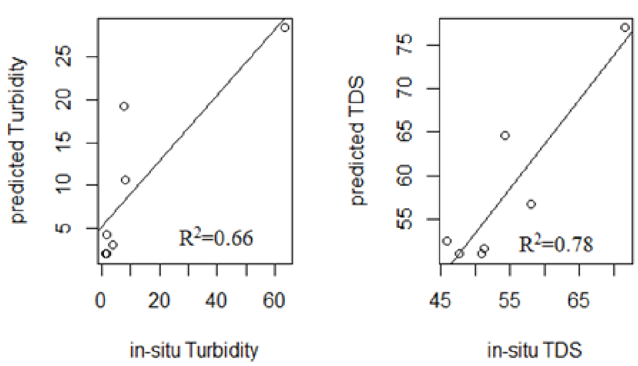

Figure 5: MLR model validation results

3.1.2. Artificial Neural Network Model

The model for estimating Turbidity had input layers, a band combination of optical bands (R*G*B), and a band ratio of green and near infrared (G/NIR). The model for estimating TDS had input layers, a band combination of blue, red, and near infrared (B*R*NIR), together with a normalized difference water index (NDWI). The two ANN model parameters are summarized in Table 3.

Table 3: ANN model parameters for retrieving Turbidity and TDS respectively

Layer	Type	Units	Mean Weight	Weight RMS
Turbidity				
Input	Input	2	-	-
Hidden 1	Tanh	7	0.441438	1.327155
Hidden 2	Tanh	3	-0.110532	0.817854
Hidden 3	Tanh	1	-0.445876	1.060671
Output	Linear	1	-0.630847	0.000000
TDS				
Input	Input	2	-	-
Hidden 1	Tanh	7	-0.032262	0.451694
Hidden 2	Tanh	5	0.013097	0.394285
Hidden 3	Tanh	1	0.276924	0.449770
Output	Linear	1	0.121476	0.000000

As indicated in Figure 6, the validation of these two models shows that these models have allowable accuracy and can be used for modeling Turbidity and TDS.

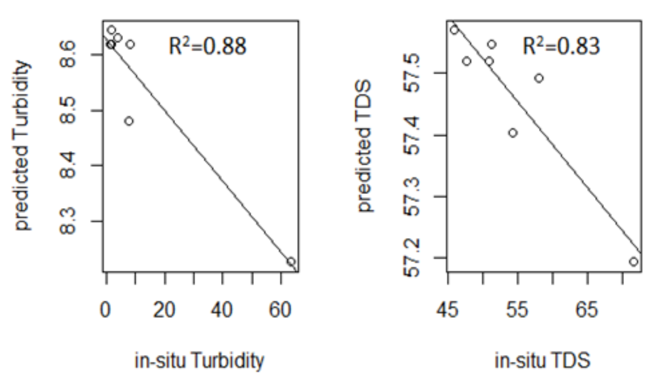

Figure 6: ANN model validation results

3.2. Comparison of ANN and MLR

From the model results, ANN performed better than MLR in estimating Turbidity and TDS. As summarized in Table 4 and Figure 7, ANN had the highest value for R^2 and the lowest standard error compared to MLR.

Table 4: MLR and ANN Models: assessment comparison results for Turbidity and TDS

	Turbidity				**TDS**			
	R^2	SE	RMSE	MAE	R^2	SE	RMSE	MAE
MLR	66.66%	0.107	14.45	7.72	78.42%	0.226	11.22	8.007
ANN	88.73%	0.0009	10.10	7.67	83.36%	0.0025	5.22	4.45

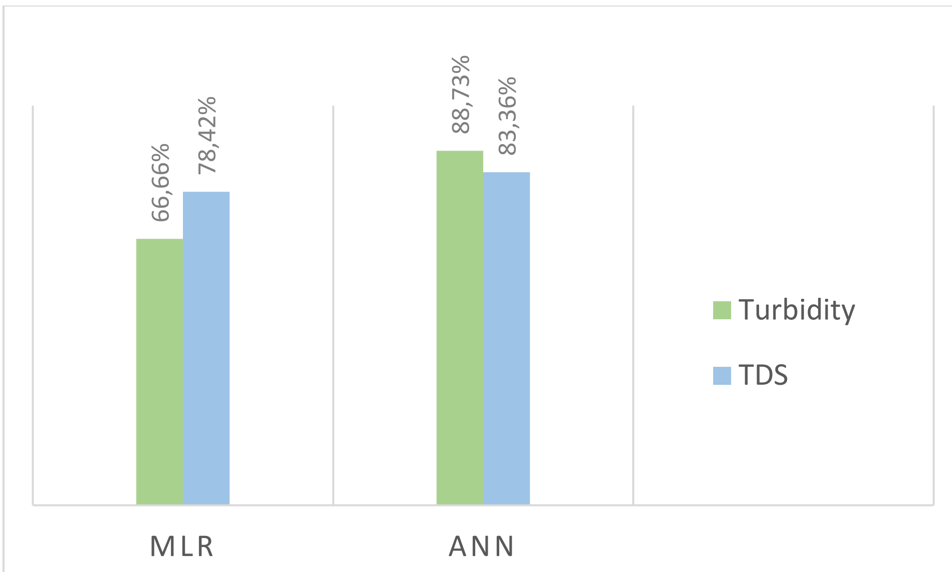

Figure 7: Goodness-of-fit between observed and remotely-sensed- water quality parameters

3.3. Modeling Turbidity and TDS

Modeling the turbidity of the waters in the study area indicated values ranging from a minimum value of 1NTU to a maximum value of 2200NTU. These results indicate that most of the area around the lake, as presented in Figure 8, has clear water (0 < 15 NTU shows clear water),. Modeling the TDS indicated TDS values ranging from the lowest value of 48mg/l to the highest of 27,000mg/l. The mean was 50.9mg/l, with a standard deviation of 83.8. As presented in the TDS distribution map in Figure 9, the values of TDS on the Mwanza coast ranged from 48mg/l to 200mg/l.

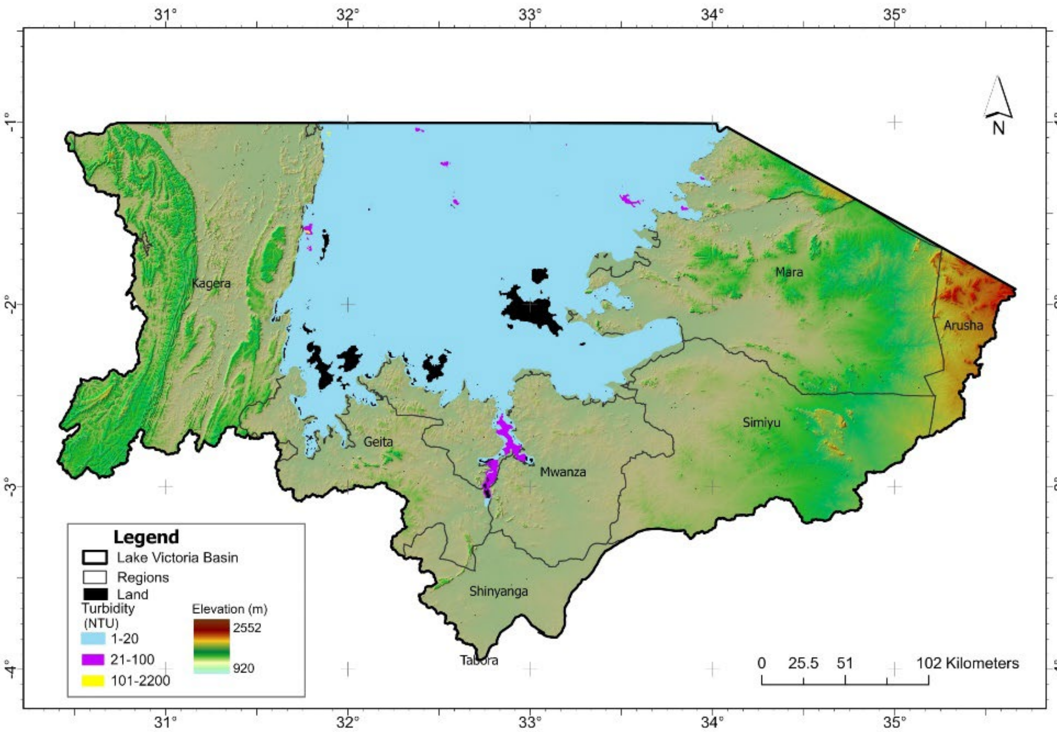

Figure 8: Turbidity distribution map

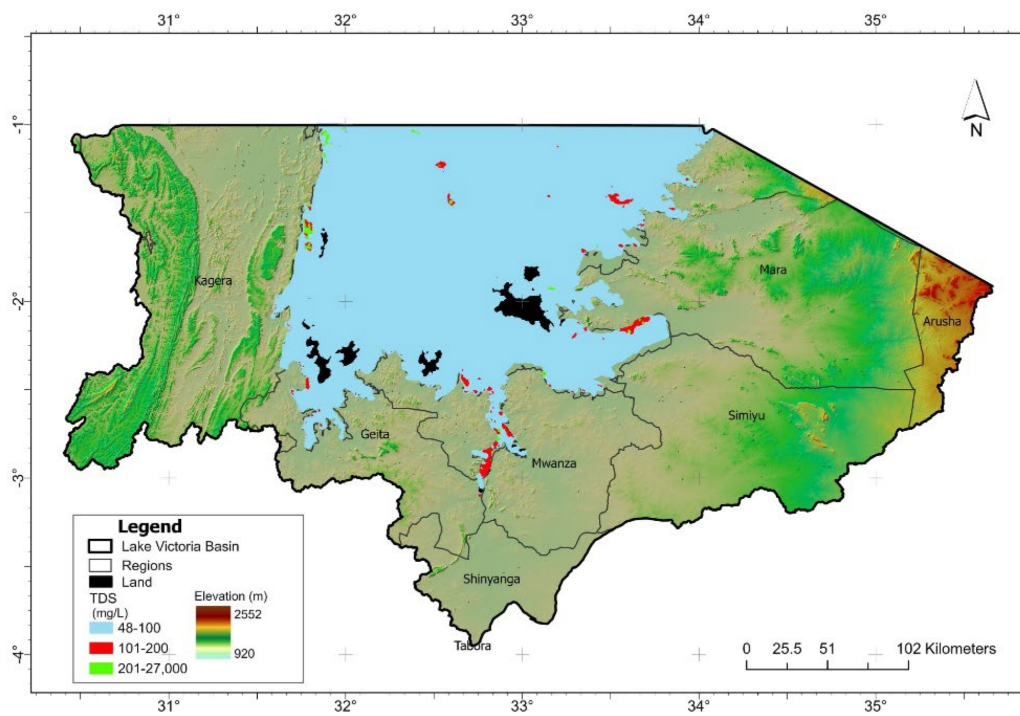

Figure 9: TDS distribution map

4. Discussion of Results

In this study, the Turbidity and TDS models were generated with the aid of MLR and ANNs (Al-Mukhtar & Al-Yaseen, 2019a). The root means square error (RMSE), standard error (SE), mean absolute error (MAE), and coefficient of determination (R^2) were used to pinpoint the accuracy of the ANN and MLR models. Figures 4 and 5 depict the validation results for the MLR and ANN modelss for both Turbidity and TDS. The results clearly show that the retrieval of Turbidity and TDS from Landsat 8 OLI using MLR and ANNs is reliable. Palabiyik and Akkan (2024) detected that MLR model can be used to predict the water quality index with very high accuracy (Palabıyık & Akkan, 2024b). However in this study, Table 5 and Figure 7 indicate that ANN is more effective than MLR in retrieving Turbidity and TDS data. ANN outperforms MLR in terms of the four criteria usedto assess model performance, namely R^2, RMSE, SE, and MAE (Abyaneh, 2014; Hafeez *et al*., 2019). This is due to the fact that ANN takes into consideration both linear and non-linear relationships between the independent and dependent variables. Additionally, the ANN models can deal with different modeling problems in rivers, lakes, and reservoirs (Y. Chen *et al*., 2020). Chen and Liu (2015) compared two ANN and MLR models in Taiwan water reservoirs and found that ANN performed better than MLR (W.-B. Chen & Liu, 2015). Tsakiri *et al.* (2018) compared these two models in predicting floods in Mohawk River, New York: they found that ANN outperformed MLR when they used a hybrid model, which improved the performance of the model. The hybrid model involved time series decomposition and artificial neural networks (Tsakiri *et al*., 2018).

The developed models can be used to monitor the quality of water in the lake by taking advantage of temporal data provided by satellite imagery. From these results, the authority can be assured that the water from the lake is safe and can be used for various purposes, such as for drinking water, irrigation, and recreational activities, without causing any health problems.

4.1. 4.1 Limitation of the developed models

The current study used water quality data acquired by the Mwanza Urban Water Supply and Sanitation Authority (MWAUWASA) from the 26 selected sampling stations (Figure 1). MWAUWASA take routine measurements of the water quality variables for the purpose of water quality monitoring. Because the observed water quality data used in this study were not normally distributed a log transformation was applied to normalize the data. This implies that the developed models might have limitations in terms of their accuracy. Further studies may be conducted using normalized observed data for better evaluation of the machine learning and regression models in estimating water quality in other parts of Lake Victoria and in other waterbodies, especially in dams and lakes across East Africa and Africa at large.

5. Conclusion and Recommendations

The main objective of this research was to compare the capability of MLR and ANN in retrieving Turbidity and Total Dissolved Solids data by using Landsat-8 OLI data. The findings indicate that Landsat 8 OLI satellite images are suitable for water quality modeling. Model evaluation results show that the ANN model outperformed MLR in retrieving Turbidity and TDS data. After mapping the distribution of turbidity and TDS levels within the lake, it was s observed that the quality of the water in Lake Vicroria is good and that it can be used for drinking purposes without causing any health problems as per the World Health Organization Guidelines on Water Quality (WHO, 2011).

Locally, the findings of this study provide essential insights for water quality and will contribute to the sustainable monitoring, assessment, and management of surface water resources. Based on the findings and conclusion of this research, it is recommended that further study of various waterbodies and an application of higher resolution satellite images be pursued. This lies in the fact that waterbodies have different characteristics based on the constituents of the waterbody which reflect differently. By testing these models in different waterbodies, it will be possible to find a universal model that can be applied in any waterbody to retrieve the associated water quality parameters. Considering the results attained for Turbidity and TDS in this study, we further recommend the application of the ANN model and Landsat 8 OLI satellite images for water quality parameter modeling.

6. Acknowledgments

The authors would like to thank the Ministry of Water and the Mwanza Urban Water Authority for providing the in-situ data and background information about the Lake Victoria water ecosystem. Landsat 8 OLI satellite images were downloaded from USGS.

7. Data Availability Statement

The remote sensing data that support the findings of this study are available from USGS website (https://earthexplorer.usgs.gov/. Upon reasonable request, the raw (*in-situ*) data that were used are available from the corresponding author.

8. References

Abyaneh, H. Z. (2014). Evaluation of multivariate linear regression and artificial neural networks in prediction of water quality parameters. *Journal of Environmental Health Science and Engineering*, *12*(1). https://doi.org/10.1186/2052-336X-12-40

Akbar, T., Hassan, Q., & Achari, G. (2010). A remote sensing-based framework for predicting water quality of different source waters. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *34*(2004), 1–4.

Ali, A., Zhou, G., Pablo Antezana Lopez, F., Xu, C., Jing, G., & Tan, Y. (2024). Deep learning for water quality multivariate assessment in inland water across China. *International Journal of Applied Earth Observation and Geoinformation*, *133*. https://doi.org/10.1016/j.jag.2024.104078

Al-Mukhtar, M., & Al-Yaseen, F. (2019a). Modeling Water Quality Parameters using Data-driven Models: a Case Study of Abu-Ziriq Marsh in the south of Iraq. *Hydrology*, *6*(1). https://doi.org/10.3390/hydrology6010024

Al-Mukhtar, M., Al-Yaseen, F., & Sahib, J. (2020). Modelling water quantity parameters using Artificial Intelligence techniques" a Case Study of Abu-Ziriq Marsh in the south of Iraq. *IOP Conference Series: Materials Science and Engineering*, *737*(1). https://doi.org/10.1088/1757-899X/737/1/012156

Arias-Rodriguez, L. F., Tüzün, U. F., Duan, Z., Huang, J., Tuo, Y., & Disse, M. (2023). Global Water Quality of Inland Waters with Harmonized Landsat-8 and Sentinel-2 using Cloud-computed Machine Learning. *Remote Sensing*, *15*(5). https://doi.org/10.3390/rs15051390

Barrett, D. C., & Frazier, A. E. (2016). Automated method for monitoring water quality using Landsat imagery. *Water (Switzerland)*, *8*(6). https://doi.org/10.3390/W8060257

Batina, A., & Krtalić, A. (2024). Integrating Remote Sensing Methods for monitoring Lake Water Quality: a Comprehensive Review. In *Hydrology,10(7) 11*,(7). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/hydrology11070092

Bhatti, A. M. (2008). *Modelling and monitoring of suspended matter in surface waters using remotely sensed data*. 165.

Bonansea, M., Ledesma, M., Rodriguez, C., & Pinotti, L. (2019). Using new remote sensing satellites for assessing water quality in a reservoir. *Hydrological Sciences Journal*, *64*(1), 34–44. https://doi.org/10.1080/02626667.2018.1552001

Chan, J. Y. Le, Leow, S. M. H., Bea, K. T., Cheng, W. K., Phoong, S. W., Hong, Z. W., & Chen, Y. L. (2022). Mitigating the Multicollinearity Problem and its Machine Learning Approach: a Review. In *Mathematics,* 10(8) *10* (8). MDPI. https://doi.org/10.3390/math10081283

Chen, W.-B., & Liu, W.-C. (2015). Water Quality Modeling in Reservoirs using Multivariate Linear Regression and Two Neural Network Models. *Advances in Artificial Neural Systems*, *2015*, 1–12. https://doi.org/10.1155/2015/521721

Chen, Y., Song, L., Liu, Y., Yang, L., & Li, D. (2020). A review of the artificial neural network models for water quality prediction. In *Applied Sciences (Switzerland), 10(17) 10*17). MDPI AG. https://doi.org/10.3390/app10175776

Ciaburro, G., & Venkateswaran, B. (2017). *Neural Networks with Remote Sensing Techniques: Smart models using CNN, RNN, deep learning, and artificial intelligence principles*. Packt Publishing.

Dai, X., Zhou, Y., Ma, W., & Zhou, L. (2017). Influence of spatial variation in land-use patterns and topography on water quality of the rivers inflowing to Fuxian Lake, a large deep lake in the plateau of southwestern China. *Ecological Engineering*, *99*, 417–428. https://doi.org/10.1016/j.ecoleng.2016.11.011

Elhag, M., Gitas, I., Othman, A., Bahrawi, J., & Gikas, P. (2019). *Assessment of Water Quality Parameters using Temporal Remote Sensing Spectral Reflectance.* In https://doi.org/10.3390/w11030556

EWURA. (2020). *Water and Wastewater Quality Monitoring Guidelines for Water Supply and Sanitation Authorities. Second Edition*. https://www.ewura.go.tz/wp-content/uploads/2020/06/Water-and-Wastewater-Quality-Monitoring-Guidelines-2020.pdf

Ewuzie, U., Aku, N. O., & Nwankpa, S. U. (2021). An appraisal of data collection, analysis, and reporting adopted for water quality assessment: A case of Nigeria water quality research. *Heliyon*, *7*(9). https://doi.org/10.1016/j.heliyon.2021.e07950

Frasson, R. P. M., Ardila, D. R., Pease, J., Hestir, E., Bright, C., Carter, N., Dekker, A. G., Thompson, D. R., Green, R. O., & Held, A. (2024). The impact of spatial resolution on inland water quality monitoring from space. *Environmental Research Communications*, *6*(10). https://doi.org/10.1088/2515-7620/ad7dd8

Gholizadeh, M. H., Melesse, A. M., & Reddi, L. (2016a). A comprehensive review on water quality parameters estimation using remote sensing techniques. In *Sensors (Switzerland),16*(8). https://doi.org/10.3390/s16081298

Gholizadeh, M. H., Melesse, A. M., & Reddi, L. (2016b). A comprehensive review on water quality parameters estimation using remote sensing techniques. In *Sensors (Switzerland),16*(8). MDPI AG. https://doi.org/10.3390/s16081298

Giardino, C., Bresciani, M., Cazzaniga, I., Schenk, K., Rieger, P., Braga, F., Matta, E., & Brando, V. E. (2014). *Evaluation of Multi-resolution Satellite Sensors for assessing Water Quality and Bottom Depth of Lake Garda*. *1*, 24116–24131. https://doi.org/10.3390/s141224116

Hafeez, S., Wong, M. S., Ho, H. C., Nazeer, M., Nichol, J., Abbas, S., Tang, D., Lee, K. H., & Pun, L. (2019). Comparison of Machine Learning Algorithms for Retrieval of Water Quality Indicators in Case-II waters: a Case Study of Hong Kong. *Remote Sensing*, *11*(6). https://doi.org/10.3390/rs11060617

Huangfu, K., Li, J., Zhang, X., Zhang, J., Cui, H., & Sun, Q. (2020). Remote estimation of water quality parameters of medium-and small-sized inland rivers using Sentinel-2 imagery. *Water (Switzerland)*, *12*(11), 1–18. https://doi.org/10.3390/w12113124

Isık, H., & Akkan, T. (2024a). Water Quality Assessment with Artificial Neural Network Models: Performance Comparison between SMN, MLP and PS-ANN Methodologies. *Arabian Journal for Science and Engineering*. https://doi.org/10.1007/s13369-024-09238-5

Isık, H., & Akkan, T. (2024b). Water Quality Assessment with Artificial Neural Network Models: Performance Comparison between SMN, MLP and PS-ANN Methodologies. *Arabian Journal for Science and Engineering*. https://doi.org/10.1007/s13369-024-09238-5

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning with Applications in R* (2nd ed.). Springer New York, NY. https://doi.org/https://doi.org/10.1007/978-1-0716-1418-1

Ji, L., Li, Y., Zhang, G., & Bi, Y. (2021). Anthropogenic disturbances have contributed to degradation of river water quality in arid areas. *Water (Switzerland)*, *13*(22). https://doi.org/10.3390/w13223305

Jin, D. (2022). Application of remote sensing technology in water quality monitoring. In *Highlights in Science, Engineering and Technology. MSESCS* (Vol. 2022).

Kapalanga, T. S., Hoko, Z., Gumindoga, W., & Chikwiramakomo, L. (2021). Remote-sensing-based algorithms for water quality monitoring in Olushandja Dam, north-central Namibia. 1878–1894. https://doi.org/10.2166/ws.2020.290

Kc, A., Chalise, A., Parajuli, D., Dhital, N., Shrestha, S., & Kandel, T. (2019). Surface Water Quality Assessment using Remote Sensing, GIS and Artificial Intelligence. In *Technical Journal, 1,* (1).

Kepford, K., & Hicks, W. (1993). Statistical Methods for the Analysis of Lake Water Quality Trends. *Technical Supplement to The Lake and Reservoir Restoration Guidance Manual*. https://nepis.epa.gov/

Kwong, I. H. Y., Wong, F. K. K., & Fung, T. (2022). Automatic Mapping and Monitoring of Marine Water Quality Parameters in Hong Kong using Sentinel-2 Image Time Series and Google Earth Engine Cloud Computing. *Frontiers in Marine Science*, *9*. https://doi.org/10.3389/fmars.2022.871470

Laili, N., Arafah, F., Jaelani, L. M., Subehi, L., Pamungkas, A., Koenhardono, E. S., & Sulisetyono, A. (2015). Development of Water Quality Parameters Retrieval Algorithms for estimating Total Suspended Solids and Chlorophyll - a Concentration using Landsat-8 Imagery: Poteran Island Water. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *2*(2W2), 55–62. https://doi.org/10.5194/isprsannals-II-2-W2-55-2015

Liu, Q., Xu, X., & Chen, M. (2024a). Changing pattern and driving factors of ecosystem service value of the lakes in Northern China since 1990. *Ecological Indicators*, *158*. https://doi.org/10.1016/j.ecolind.2023.111370

Liu, Q., Xu, X., & Chen, M. (2024b). Changing pattern and driving factors of ecosystem service value of the lakes in Northern China since 1990. *Ecological Indicators*, *158*. https://doi.org/10.1016/j.ecolind.2023.111370

Loaiza, J. G., Rangel-Peraza, J. G., Monjardín-Armenta, S. A., Bustos-Terrones, Y. A., Bandala, E. R., Sanhouse-García, A. J., & Rentería-Guevara, S. A. (2023). Surface Water Quality Assessment through Remote Sensing based on the Box–Cox Transformation and Linear Regression. *Water (Switzerland)*, *15*(14). https://doi.org/10.3390/w15142606

López, O. A. M., López, A. M., & Jose, C. (2022). Fundamentals of Artificial Neural Networks and Deep Learning. In *Multivariate Statistical Machine Learning Methods for Genomic Prediction.*379–425. Springer International Publishing. https://doi.org/10.1007/978-3-030-89010-0_10

Mishra, P., Pandey, C. M., Singh, U., Gupta, A., Sahu, C., & Keshri, A. (2019). Descriptive statistics and normality tests for statistical data. *Annals of Cardiac Anaesthesia*, *22*(1), 67–72. https://doi.org/10.4103/aca.ACA_157_18

Mohammed, H., Michel Tornyeviadzi, H., & Seidu, R. (2022). Emulating process-based water quality modelling in water source reservoirs using machine learning. *Journal of Hydrology*, *609*. https://doi.org/10.1016/j.jhydrol.2022.127675

Montgomery, D. C., Peck, A. E., & Vining, G. G. (2012). *Introduction to Linear Regression Analysis* (6th ed.). John Wiley & Sons, Inc.

Muhoyi, H., Gumindoga, W., Mhizha, A., Misi, S. N., & Nondo, N. (2022). Water quality monitoring using remote sensing: Lower Manyame Sub-catchment, Zimbabwe. *Water Practice and Technology*, *17*(6), 1347–1357. https://doi.org/10.2166/wpt.2022.061

Nakkazi, M. T., Nkwasa, A., Martínez, A. B., & van Griensven, A. (2024). Linking land use and precipitation changes to water quality changes in Lake Victoria using earth observation data. *Environmental Monitoring and Assessment*, *196*(11), 1104. https://doi.org/10.1007/s10661-024-13261-2

Ochaeta, G., Skinner-alvarado, J., Violeta, M., Moran, R., & Hanan, N. P. (2020). *Hyperspectral Satellite Remote Sensing of Water Quality in Lake*. *8* (*February*). https://doi.org/10.3389/fenvs.2020.00007

Palabıyık, S., & Akkan, T. (2024a). Evaluation of water quality based on artificial intelligence: performance of multilayer perceptron neural networks and multiple linear regression *versus* water quality indexes. *Environment, Development and Sustainability*. https://doi.org/10.1007/s10668-024-05075-6

Palabıyık, S., & Akkan, T. (2024b). Evaluation of water quality based on artificial intelligence: performance of multilayer perceptron neural networks and multiple linear regression versus water quality indexes. *Environment, Development and Sustainability*. https://doi.org/10.1007/s10668-024-05075-6

Pourghasemi, H. R., & Gokceoglu, C. (2019). *Spatial Modeling in GIS and R for Earth and Environmental Sciences*. Elsevier.

Pu, F., Ding, C., Chao, Z., Yu, Y., & Xu, X. (2019). Water-quality classification of inland lakes using Landsat8 images by convolutional neural networks. *Remote Sensing*, *11*(14). https://doi.org/10.3390/rs11141674

Ritchie, J. C., Zimba, P. V., & Everitt, J. H. (2003). Remote sensing techniques to assess water quality. In *Photogrammetric Engineering and Remote Sensing,*69(6)*69* (6), 695–704. American Society for Photogrammetry and Remote Sensing. https://doi.org/10.14358/PERS.69.6.695

Sagan, V., Peterson, K. T., Maimaitijiang, M., Sidike, P., Sloan, J., Greeling, B. A., Maalouf, S., & Adams, C. (2020). Monitoring inland water quality using remote sensing: potential and limitations of spectral indices, bio-optical simulations, machine learning, and cloud computing. *Earth Science Reviews*, *205*, 103187. https://doi.org/https://doi.org/10.1016/j.earscirev.2020.103187

Shanmuganathan, S. (2016). Artificial Neural Network Modelling: an Introduction. In S. Shanmuganathan & S. Samarasinghe (Eds.), *Artificial Neural Network Modelling* (1st ed), 1–14. Springer Cham. https://doi.org/https://doi.org/10.1007/978-3-319-28495-8

Sidabutar, N. V., Namara, I., Hartono, D. M., & Soesilo, T. E. B. (2017). The effect of anthropogenic activities on the decrease of water quality. *IOP Conference Series: Earth and Environmental Science*, *67*(1). https://doi.org/10.1088/1755-1315/67/1/012034

Srivastava, P. K., Malhi, R. K. M., Pandey, P. C., Anand, A., Singh, P., Pandey, M. K., & Gupta, A. (2020). Revisiting hyperspectral remote sensing: origin, processing, applications and way forward. In P. C. Pandey, P. K. Srivastava, H. Balzter, B. Bhattacharya, & G. P. Petropoulos (Eds.), *Hyperspectral Remote Sensing,* 3–21. Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-08-102894-0.00001-2

Thai-Hoang, L., Thong, T., Loc, H. T., Van, P. T. T., Thuy, P. T. P., & Thuoc, T. L. (2022). Influences of anthropogenic activities on water quality in the Saigon River, Ho Chi Minh City. *Journal of Water and Health*, *20*(3), 491–504. https://doi.org/10.2166/WH.2022.233

Tsakiri, K., Marsellos, A., & Kapetanakis, S. (2018). Artificial neural network and multiple linear regression for flood prediction in Mohawk River, New York. *Water (Switzerland)*, *10*(9). https://doi.org/10.3390/w10091158

Tu, M., Smith, P., & Filippi, A. M. (2018). *Hybrid forward-selection method-based water-quality estimation via combining Landsat TM , ETM + , and OLI / TIRS images and ancillary environmental data*. 1–23.

URT. (2007). *The Environmental Management (Water Quality Standards) Regulations, 2007*. https://www.nemc.or.tz/uploads/publications/sw-1660810388The%20Environmental%20Management%20(water%20Quality%20Standards)Regulations,%202007%20.pdf

URT. (2018). *National Guideline on Drinking Water Quality Monitoring and Reporting*. https://www.maji.go.tz/uploads/publications/sw1541500878-National%20Guideline%20on%20Drinking%20Water%20Quality%20Monitoring%20and%20Reporting.pdf

World Health Organization. (2011). *Guidelines for drinking-water quality: Fourth edition incorporating the first and second addenda.* https://iris.who.int/bitstream/handle/10665/44584/9789241548151_eng.pdf

Yang, G., & Moyer, D. L. (2020). Estimation of nonlinear water-quality trends in high-frequency monitoring data. *Science of the Total Environment*, *715*. https://doi.org/10.1016/j.scitotenv.2020.136686

Yang, H., Kong, J., Hu, H., Du, Y., Gao, M., & Chen, F. (2022). A Review of Remote Sensing for Water Quality Retrieval: Progress and Challenges. *Remote Sensing*, *14*(8), 1770. https://doi.org/10.3390/rs14081770

Zhu, X., Wang, L., Zhang, X., He, M., Wang, D., Ren, Y., Yao, H., Ngegla, J. V., Pan, H. . (2022). Effects of different types of anthropogenic disturbances and natural wetlands on water quality and microbial communities in a typical black-odor river. *Ecological Indicators*, *136*. https://doi.org/10.1016/j.ecolind.2022.108613