

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/341548423>

Use of Artificial Neural Networks and Multiple Linear Regression Model for the Prediction of Dissolved Oxygen in Rivers: Case Study of Hydrographic Basin of River Nyando, Kenya

Article in *Complexity* · May 2020

DOI: 10.1155/2020/9570789

CITATIONS

15

READS

191

3 authors, including:



Yashon Ouma

Moi University

37 PUBLICATIONS 1,088 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Water Quality Modelling using Multisensory Data with Artificial Intelligence and Empirical Regression [View project](#)



Mobile Mapping Systems for Pavement Condition Mapping and Evaluation. AvH Germany/HFT Stuttgart [View project](#)

Research Article

Use of Artificial Neural Networks and Multiple Linear Regression Model for the Prediction of Dissolved Oxygen in Rivers: Case Study of Hydrographic Basin of River Nyando, Kenya

Yashon O. Ouma ^{1,2}, Clinton O. Okuku,¹ and Evalyne N. Njau¹

¹Department of Civil and Structural Engineering, Moi University, Eldoret 30100, Kenya

²Department of Civil Engineering, University of Botswana, Private Bag UB 0061, Gaborone, Botswana

Correspondence should be addressed to Yashon O. Ouma; yashon_o@hotmail.com

Received 15 November 2019; Revised 7 April 2020; Accepted 5 May 2020; Published 20 May 2020

Academic Editor: Ana Meštrović

Copyright © 2020 Yashon O. Ouma et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The process of predicting water quality over a catchment area is complex due to the inherently nonlinear interactions between the water quality parameters and their temporal and spatial variability. The empirical, conceptual, and physical distributed models for the simulation of hydrological interactions may not adequately represent the nonlinear dynamics in the process of water quality monitoring and prediction, especially in watersheds with scarce water quality monitoring networks. To overcome the lack of data in water quality monitoring and prediction, this paper presents an approach based on the feedforward neural network (FNN) model for the simulation and prediction of dissolved oxygen (DO) in the Nyando River basin in Kenya. To understand the influence of the contributing factors to the DO variations, the model considered the inputs from the available water quality parameters (WQPs) including discharge, electrical conductivity (EC), pH, turbidity, temperature, total phosphates (TPs), and total nitrates (TNs) as the basin land-use and land-cover (LULC) percentages. The performance of the FNN model is compared with the multiple linear regression (MLR) model. For both FNN and MLR models, the use of the eight water quality parameters yielded the best DO prediction results with respective Pearson correlation coefficient R values of 0.8546 and 0.6199. In the model optimization, EC, TP, TN, pH, and temperature were most significant contributing water quality parameters with 85.5% in DO prediction. For both models, LULC gave the best results with successful prediction of DO at nearly 98% degree of accuracy, with the combination of LULC and the water quality parameters presenting the same degree of accuracy for both FNN and MLR models.

1. Introduction

Increased surface water pollution due to urbanization, excessive water consumptions, population growth, industrial wastewater discharge, and agricultural activities results in low dissolved oxygen (DO) levels and worsens the existence conditions in aquatic systems [1–4]. Quantification of dissolved oxygen is thus important for evaluating surface water quality because of its representation of the level of pollution and the state of aquatic ecosystem [5–7]. For near-hypoxic river systems, like Nyando River in Kenya, the accurate prediction of DO levels remains a challenge due to lack of sufficient and accurate water quality monitoring networks.

The process of predicting water quality over a catchment is complex and nonlinear and exhibits both temporal and spatial variability [8, 9]. The models developed to simulate the process can be categorized as empirical, conceptual, and physically based distributed models. Although parametric statistical and deterministic models have been the traditional approaches for modeling water quality, these models require vast information and data on various hydrological sub-processes in order to arrive at the end results [10–12]. Moreover, these models require precisely determined rate constants and coefficients pertaining to various hydrological, chemical, physical, and biological processes, which are largely time and space specific in nature. Additionally,

though these models have analytical solutions, they have boundary conditions as limitations [13–15].

In order to overcome these limitations, the use of knowledge-based systems, genetic algorithms, artificial neural networks, and fuzzy inference systems for modeling water quality parameters has been proposed [16, 17]. Because of the ability to learn the temporal dynamics of a system with less input data and efficiency in solving nonlinear problems, different artificial neural networks (ANNs) have been tested for water quality prediction [17]. Several studies have used different ANN architectures for the prediction of DO in different case studies using varied estimation parameters. For example, to simulate the DO concentrations in Surma River in Bangladesh, Ahmed [18] used BOD and COD data collected over a period of 3 years. The study presented and compared the simulation models based on feedforward neural network (FFNN) and radial basis function neural network (RBFNN), with RBFNN outperforming the FFNN with Pearson correlation coefficient R values of 0.96, as compared to 0.904 for model testing and validation, respectively. Singh et al. [10] and Dogan et al. [19] also used feedforward NN models for the computation of dissolved oxygen and BOD and dissolved oxygen, respectively, for river waters. Palani et al. [20] demonstrated the application of neural network models for the prediction and forecasting of selected water quality variables.

Abba et al. [21] used monthly data for the period of 1999–2005 to predict DO in downstream of Yamuna River in Agra city in India using the input variables of DO, pH, BOD, and water temperature. The study compared experimental multilinear regression (MLR), adaptive neuro fuzzy inference system (ANFIS), and ANN models for the prediction of DO. By varying the input parameters, ANN was a better predictor for the DO with up to 94% accuracy as compared to ANFIS and MLR with average of 81% accuracy. In Nzoia River in Lake Victoria basin (Kenya), Kanda et al. [22] used monthly data from 2003 to 2013 comprising of pH, turbidity, temperature, and electrical conductivity to predict DO using multilayer perception (MLP), a form of feedforward backpropagation ANN. The number of input neurons varied from 1 to 4 representing input parameters that affect DO, and the number of neurons in hidden layer varied from 22 to 28. In contrast to the results in [21], the exclusion of pH exhibited acceptable results in the prediction of DO. Sarkar and Pandey [23], using the feedforward back propagation network architecture, predicted the DO using discharge, temperature, pH, BOD, and COD as the input variables for River Yamuna in India. The study concluded that the performance of the ANN model is the best with optimum input variables, and the values above the optimum would cause the model to overfit data, while values below the optimum would result in inaccurate prediction. Several other studies (e.g., [24] and [25]) also used different water quality parameters with different ANN architecture models to simulate and predict DO. Because of lack of water quality monitoring data and the nonlinear and complex nature of interactions between water quality parameters, ANN has been proposed for the simulation of water quality parameters [26–29].

While several studies have used different ANN architectures to model different water quality parameters within river catchment systems (e.g., [10, 30–33]), land-use and land-cover (LULC) information, which directly affects river water quality by altering sediment, chemical loads, and watershed hydrology, has been largely omitted in the modeling process [34]. The effects of LULC on water quality and quantity can be explored through various techniques varying from regression-based methods such as linear and multilinear regression to watershed models. In understanding the relationship between LULC and river water quality, several studies have been carried out. For example, Bonansea et al. [1] considered the water quality parameters such as temperature, pH, DO, discharge, total phosphorus, and total nitrogen and the influence of LULC categories including bare land, gravels, bare ground, and bare rocks, natural forest including natural shrub, thicket and herb, and timber plantation, agriculture including agricultural and livestock developments, and urban including rivers, reservoirs, wetland, and sandy beach on the water quality in Rio Tercero, Argentina. The study indicated significant influence of land use on the water quality parameters. Overall, the increase in agricultural activities and urban developments was observed to be responsible for decline in water quality.

Kalin and Isik [35] conducted studies on the impacts of LULC on water quality in 18 watersheds in West Georgia. Using a three-layer feedforward ANN with input variables comprising of LULC percentages, streamflow, and temperature, the study concluded that LULC affects water quality by altering sediment, watershed hydrology, and chemical loads. Ahearn et al. [36] used linear mixed effects (LME) to establish the relationship between LULC and the water quality parameters and concluded that agricultural coverage had the most significant effect on water quality within the Cosumnes watershed, particularly on total suspended solids (TSSs) and nitrate concentration. From previous research, very few studies have tried to incorporate water quality parameters and LULC in the prediction of the variability of DO in river water systems [37].

The Nyando River is a major river system across the rural parts of western Kenya. The water quality of the river is continuously degrading due to the amounts of effluents discharged into the river system from industrial and agricultural wastewater. Because of this, several portions of the river are considered as near-hypoxic systems. There is thus the need to develop a case-study model for the basin to inform on identification of efficient management strategies. Because of the limited water quality data and to overcome the difficulties in DO prediction in near-hypoxic river systems, this study proposes the use of feedforward neural network (FNN) for the prediction of DO in the Nyando River basin. The advantage of FNN is that even with a single hidden layer and arbitrary bounded and smooth activation function, the network is capable of approximating a continuous nonlinear function. Further, FNN has no *a priori* assumptions about the relationships between the independent and dependent variables. The performance of the nonparametric ANN model is compared with the parametric multiple linear regression (MLR) models. MLR is an

established statistical method suitable for the establishment of the linear relationships between input-output data variables for intercomparison of models.

The aim of this study is to design a feedforward neural network model for the prediction of dissolved oxygen concentrations in river waters within the Nyando River basin and to demonstrate its application in identifying the complex nonlinear relationships between input water quality parameters and LULC. The water quality data used in the study include temperature, pH, discharge, turbidity, total suspended solids (TSSs), electrical conductivity (EC), total phosphates (TPs), and total nitrates (TNs). The data were collected from 2006–2011. It is expected that the proposed approach for dissolved oxygen prediction enables (a) the selection of optimal water quality variables for predicting DO; (b) integration of water quality variables with LULC; and (c) prediction of DO concentrations from sparsely available data.

2. Materials and Methods

2.1. Characterization of the Study Area. Nyando River basin is one of the seven major river basins in Kenya, and it covers an area of approximately 3,550 km². It is bounded by latitudes 0°7'48"N and 0°24'36"S and longitudes 34°24'36"E and 35°43'12" (Figure 1). Nyando River drains into Lake Victoria, at altitudes of about 1,300 m above mean sea level. The regional climate is influenced by the Equatorial Convergence Zone (ITCZ) and modified by orographic effects. Land-use and land-cover types vary from forests in uplands to mixed-type subsistent agriculture in the mid to lowland parts. The human population is about 800,000 people, and it is greatly responsible for the LULC changes within the basin. The climate of Nyando basin varies from subhumid to humid, due to the variation in altitude from the highlands to the shores of Lake Victoria. The mean annual rainfall of the basin varies from 1,000 mm in regions near Lake Victoria to 1,600 mm in the highlands, and the basin is characterized by ferralsols, nitisols, cambisols, and acrisols as the main soil types [38].

The slope inclination of the highlands is 14–30%, 7–13% for the midlands, and 0–6% for the lowlands. 58% of the Nyando catchment is lowland areas, 18% highland area, and 23% for the midland area. Due to the steep topography of the basin, there is occurrence of erosion at various locations of the Nyando River. In Figure 1, the sub-basins are identified with numbers from 1 to 13. Notably, the quality of river waters within catchments is influenced by anthropogenic activities and natural processes as they interfere with water quality and impair their use for domestic, industrial, agricultural, or other purposes. Nyando River basin, like most river catchments in several developing countries, hardly has continuous and integrated river water quality monitoring networks.

2.2. Water Quality Parameters. The water quality parameters were obtained from Lake Victoria South Water Service Board (LVSWSB, Kisumu), for eight stations from

2006–2011 with their spatial location distributions shown in Figure 1. The water sampling and testing are carried out on a monthly frequency. In this study, the mean annual water quality parameters were used. The parameters analyzed in this paper are based on variables that reflect the water quality and are divided into the following categories: water quality variables indicative of stratification comprising of temperature, dissolved oxygen, and pH; water quality variables that indicate the trophic status characterized by TP, turbidity, and TN; and mineral budget which comprised of electrical conductivity (EC), TSS, and flow characterized by discharge. The annual averages for the temporal variations of the water quality parameters as measured from 2006–2011 are presented in Figure 2.

Apart from the variability of the water quality parameters and the river discharge (Figure 2(a)–2(h)), the dissolved oxygen is observed to vary consistently from station to station during the study period of 2006–2011 (Figure 2(i)). The spatial distribution of the DO concentration within the basin as interpolated using ordinary Kriging is presented in Figure 2(j). Regions with higher DO concentrations are around the intensive agricultural activities. The variability in DO is dependent on the input parameters including the water quality parameters and land use/land cover. In this study, WQP and LULC are considered as independent parameters since they are derived from the hydrologically independent sub-basins.

In Figures 1 and 2, the station IDs and the corresponding stations names are as follows: 1GBO3—Ainamutua at Kibigori; 1GB11—Ainapsiwa; 1GD03—Nyando; 1GE01—Awach Kano; 1HA11—Nyamasaria; 1GC03—Kipchorian; 1GC01—Namuting; and 1GD07—Nyando at Muhoroni.

2.3. Land Use and Land Cover (LULC) within the Nyando River Basin. Using maximum-likelihood supervised classification of 30-meter spatial resolution Landsat ETM+, the LULC of the basin was derived. By classifying and comparing the LULC changes from 2006–2008, 2008–2011, and 2006–2011, it was observed that there was insignificant difference between the LULC for the 5 years from 2006–2011. The LULC for the year 2008 was chosen as it was midway between 2006 and 2011. Figure 3 presents the LULC and the stream network within the Nyando River basin. The basin was divided into eight sub-basins according to the pour points, which correspond to the sampling stations as spatially represented in Figure 1. The rationale for the division of the basin into its sub-basins is to be able to model and understand the impacts of LULC activities and the water quality parameters on the DO within the independent sub-basins. The derived LULC in percentages for the eight sub-basins are presented in Table 1, and the classification results show that the entire basin comprised of forest (16.8%), wetlands (0.5%), shrubs (2%), and agricultural land (80.7%). The classification results show that most of the studied stations are located in agricultural land areas, though the upstream areas are mostly characterized by forests and light vegetation cover (Figure 3).

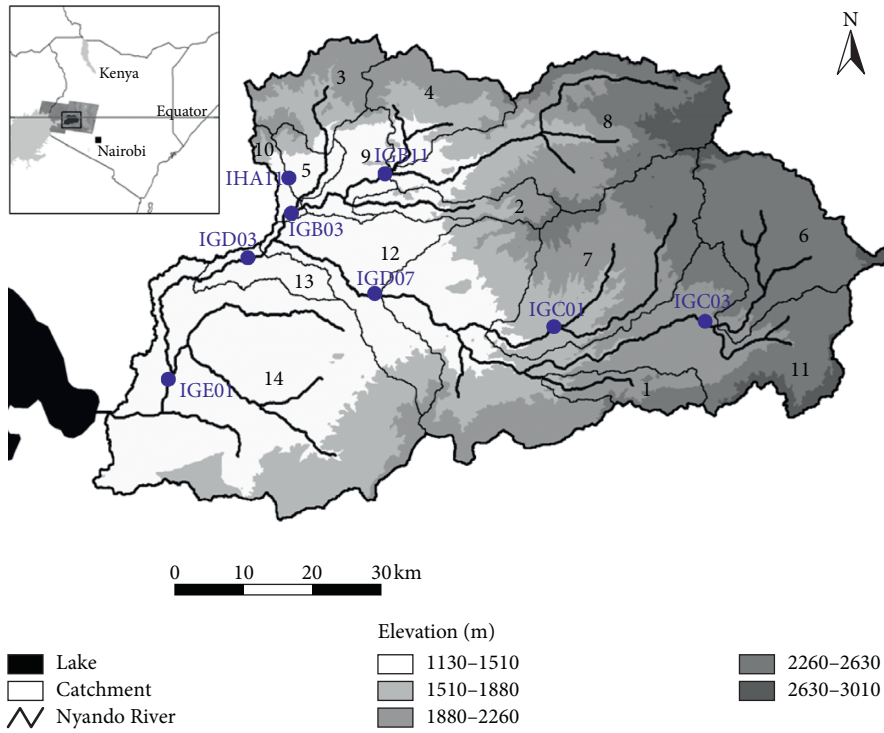
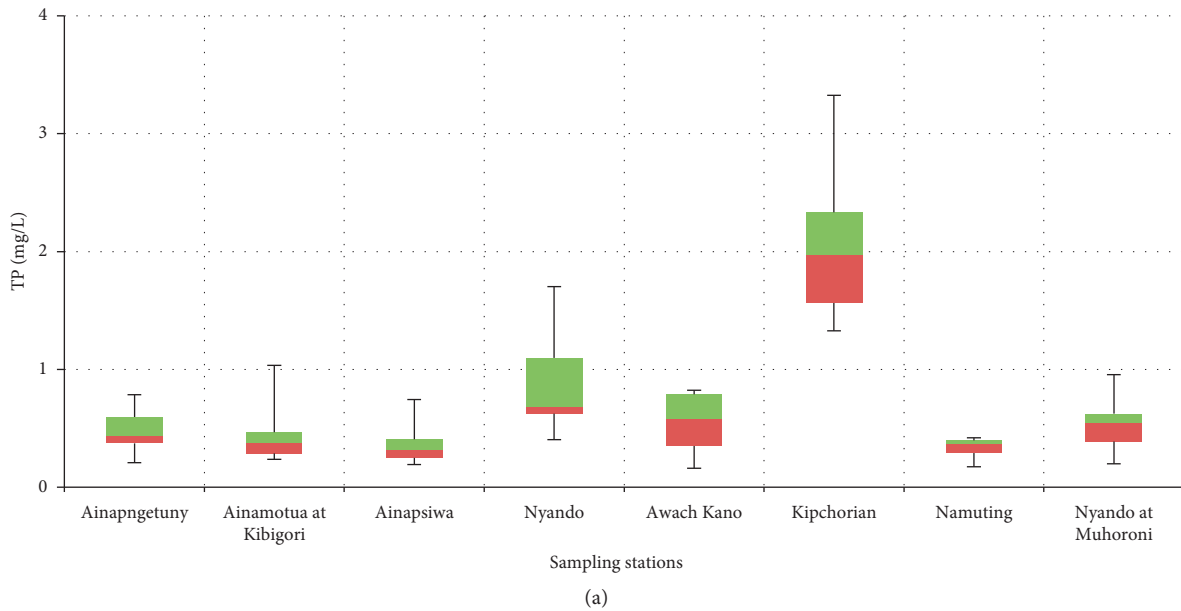
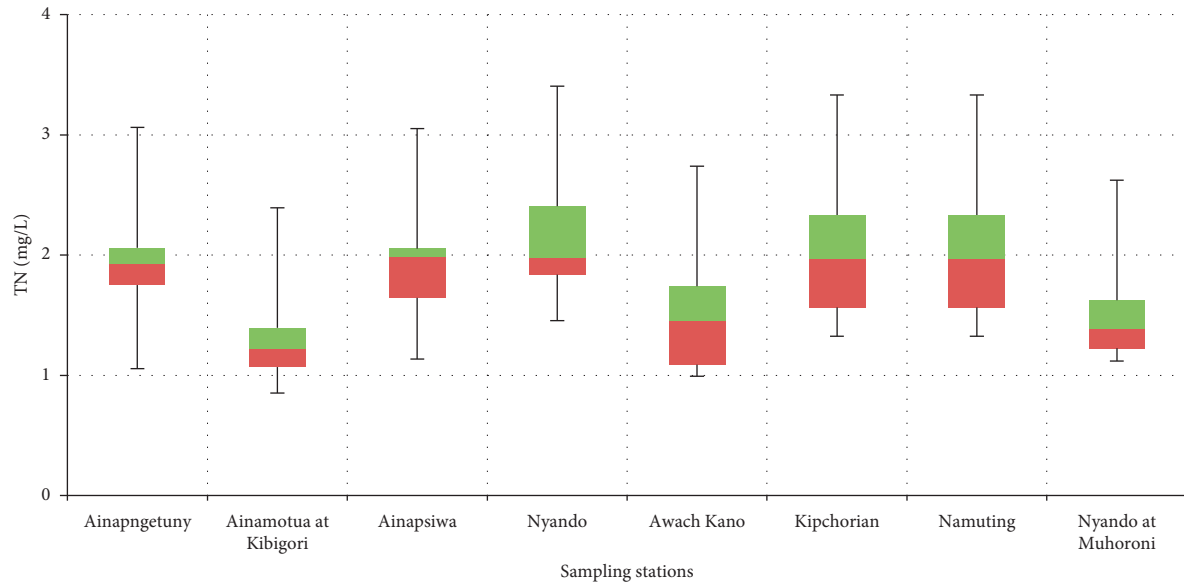


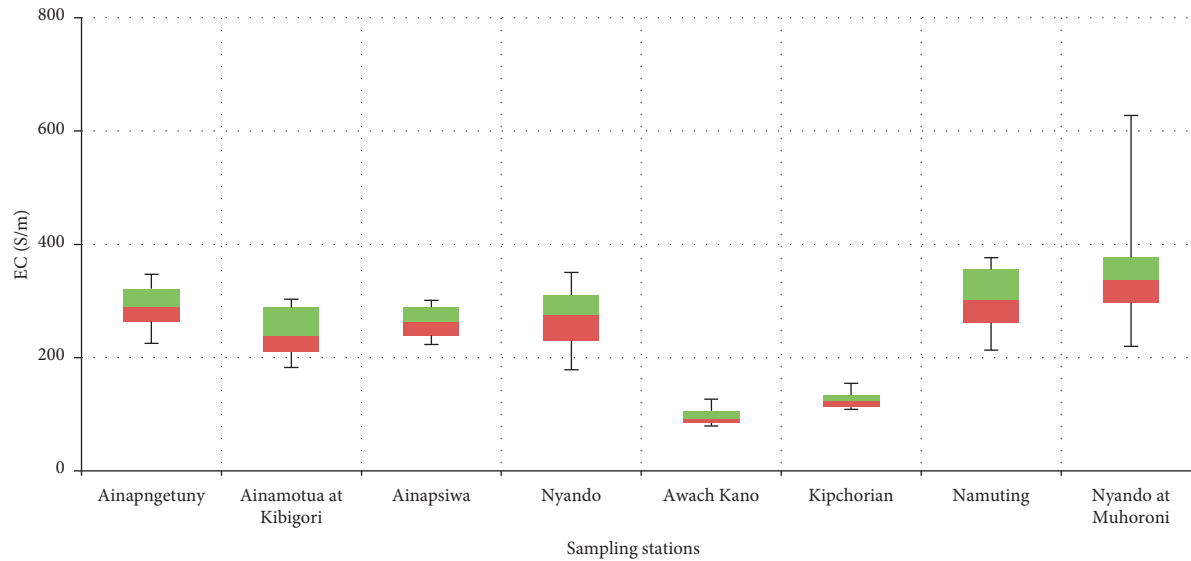
FIGURE 1: Location map of Nyando River basin [38], the sub-basins (1-13), and study sampling stations.



(a)
FIGURE 2: Continued.

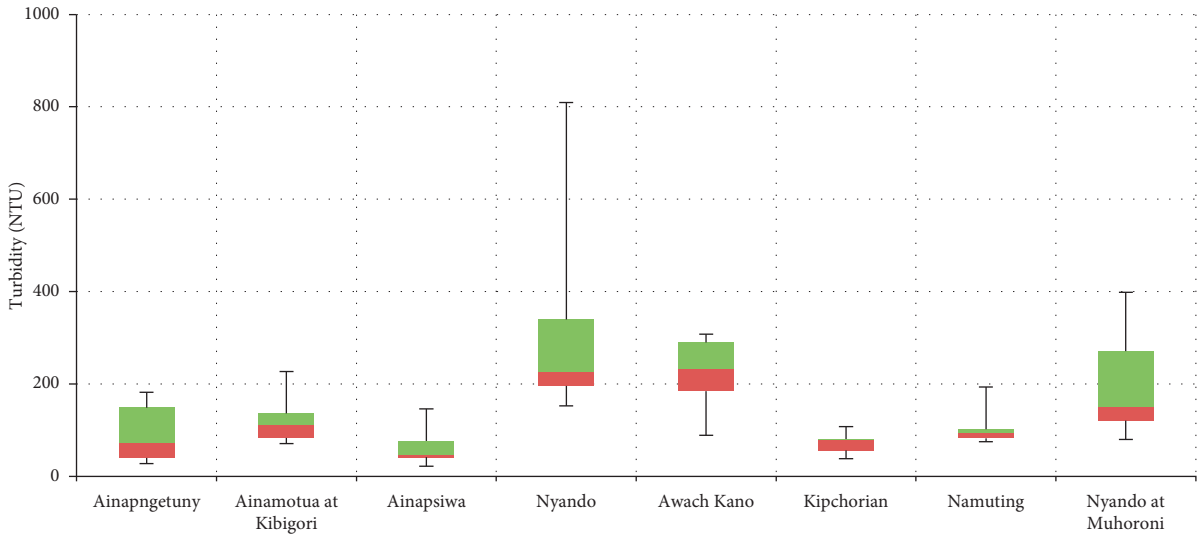


(b)

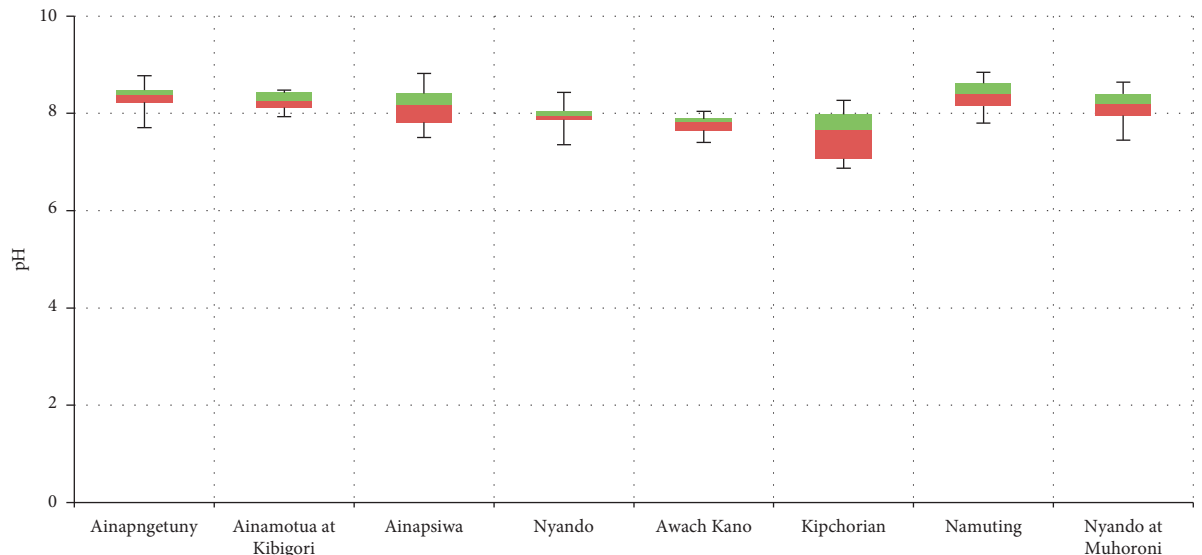


(c)

FIGURE 2: Continued.



Sampling stations
(d)



Sampling stations
(e)

FIGURE 2: Continued.

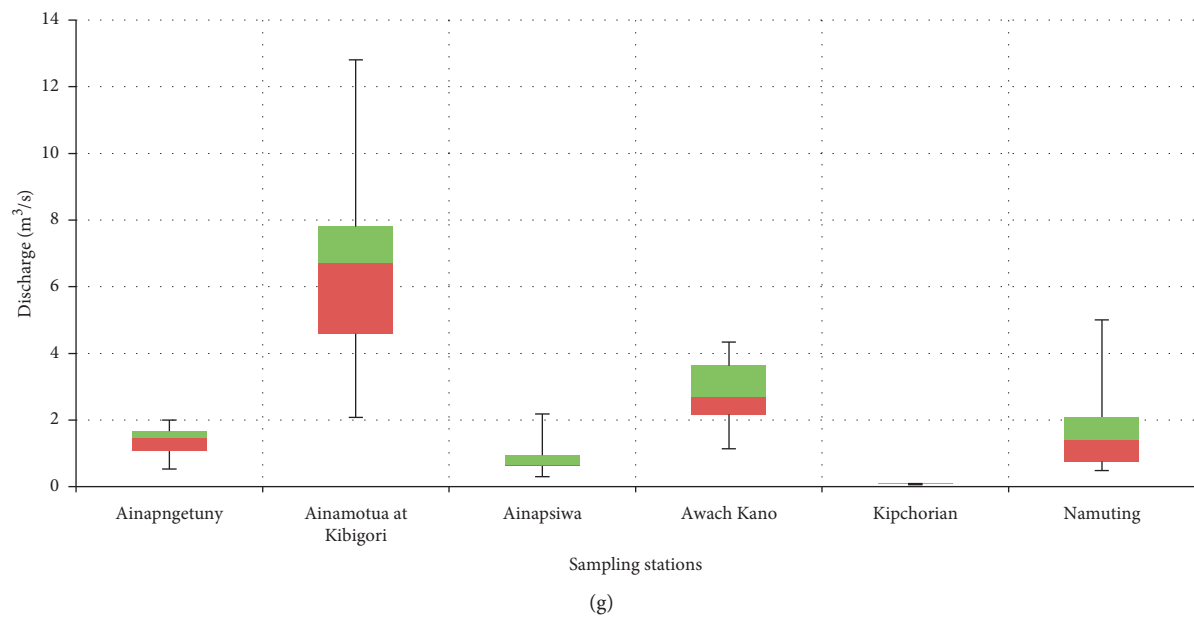
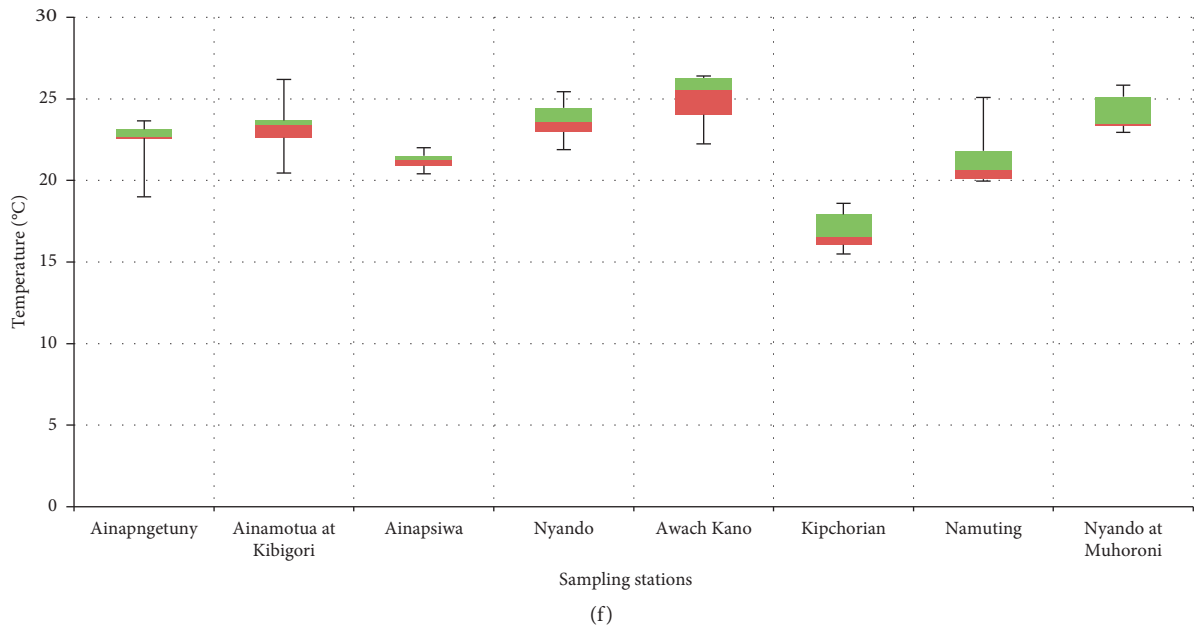
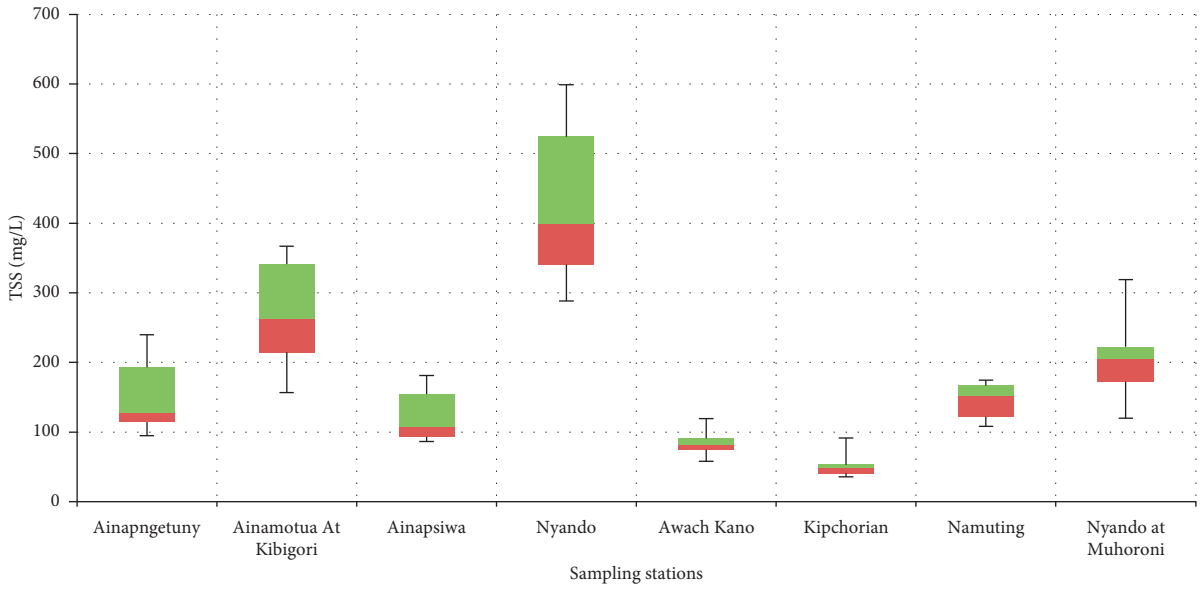
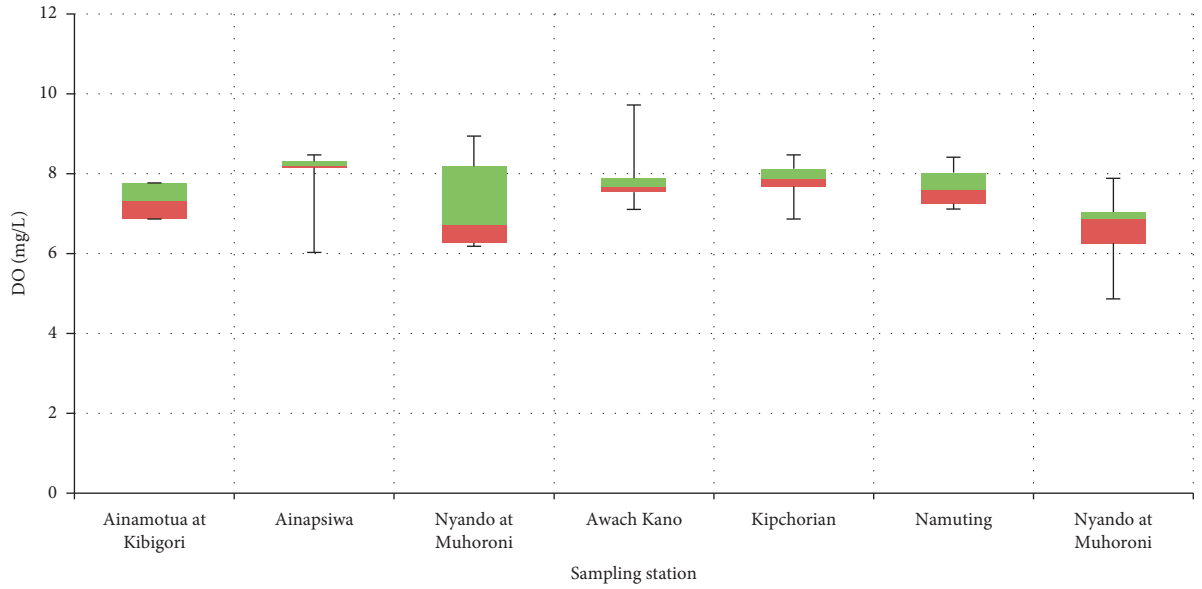


FIGURE 2: Continued.



(h)



(i)

FIGURE 2: Continued.

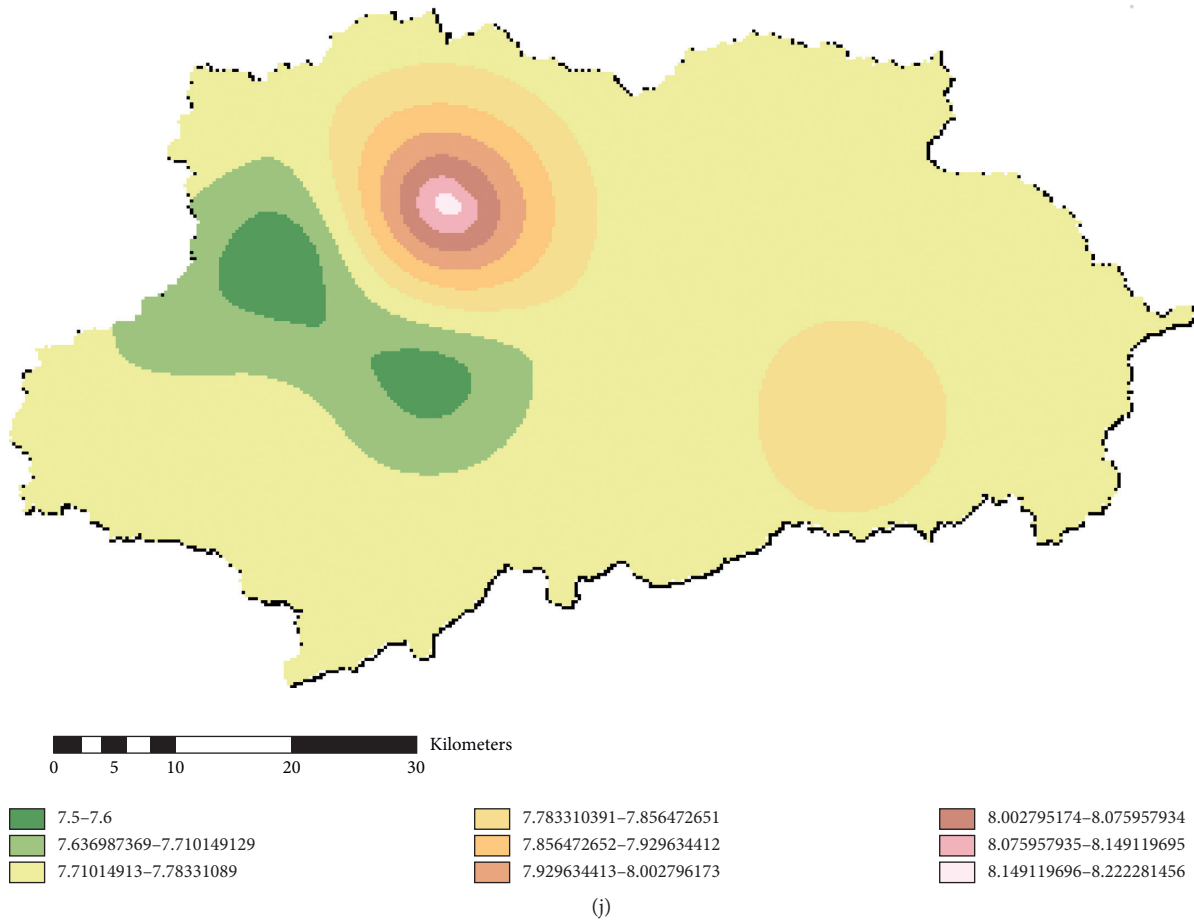


FIGURE 2: Mean annual variations of water quality parameters in the Nyando River basin from 2006–2011: (a) total phosphates; (b) total nitrates; (c) electrical conductivity; (d) turbidity; (e) pH; (f) temperature; (g) discharge; (h) total suspended solids (TSSs); (i) dissolved oxygen (the box plots denote 25th, 50th, and 75th percentiles and the whiskers indicate the 5th and 95th percentiles). (j) Spatial distribution of DO concentration within the basin.

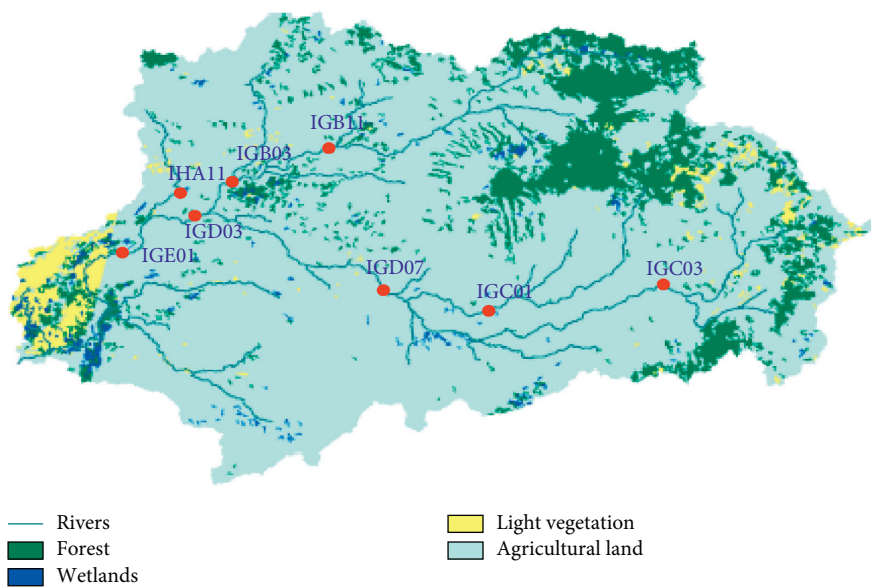


FIGURE 3: LULC of Nyando River basin in 2008 and the sampling stations.

TABLE 1: LULC in percent for the eight (8) sub-basins corresponding to the studied stations.

Watershed ID (pourpoint)	Forest (%)	Wetlands (%)	Low vegetation (shrubs) (%)	Agricultural land (%)
1GE01	16.18	0.86	1.67	81.29
1GC01	16.22	0.83	1.78	81.50
1GD01	8.50	0.17	3.16	88.17
1GD03	14.97	0.66	2.10	82.27
1GB03	9.88	0.22	0.93	88.97
1GB11	24.23	0.46	0.53	74.78
1GB05	21.06	0.30	4.74	73.91
1GC06	23.07	1.22	0.69	75.03

2.4. Multiple Linear Regression Model. Regression models are suitable for investigating the existing relationships between dependent and independent variables especially in small sample sizes [39], based on least squares fitting. In this study, the parametric MLR model is used to model the relationship between DO, the water quality variables, and LULC parameters as a linear function [40]. MLR as a parametric statistical model assumes that the variable components are independent and may not match the actual situation. The best MLR formulation is based on the highest multiple correlation coefficient (R), the lowest standard deviation, and the magnitude of the F -ratio and can also reveal the statistically significant variables of the system [41]. The general MLR model is expressed as follows:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_q x_{qi} + s_i, \quad (1)$$

where y_i is the dependent variable observed values; n is the sample size and $i = 1, \dots, n$; x_1, x_2, \dots, x_q are the explanatory or independent variables; $x_{1i}, x_{2i}, \dots, x_{qi}$ are the descriptors of observed values; s_i is the residual or error for individual i ; β_0 is a constant; and $\beta_1, \beta_2, \dots, \beta_q$ are the multiple regression coefficients [42].

In equation (1), Y represents the concentration of dissolved oxygen (DO) as the dependent variable, and (x_1, x_2, \dots, x_p) is the set of predictor p variables comprising of temperature, pH, discharge, electrical conductivity, total phosphorus, total nitrogen, and turbidity. In the implementation of stepwise regression model, MLR analysis was performed to estimate the DO and to yield variable F -significance probability. The optimal model through regression statistical feature value (P , R^2) was validated using the standard value 0.05 and excluded value 0.10.

2.5. ANN and Training Algorithm. Artificial neural networks are made up of a set of simple elements; the artificial neurons are motivated by the biological nervous systems. There are different architectures and models for ANN, namely, multilayer perceptron (MLP), adaptive neuro fuzzy inference system (ANFIS), recurrent neural network (RNN), generalized regression neural network (GRNN), and radial basis function network (RBFN). These ANN models can be categorized into feedforward neural networks and recurrent neural networks. MLP neural networks, trained with a backpropagation learning algorithm, are the most popular feedforward neural networks (FNNs) and have been widely

used in hydrologic forecasting models (e.g., [43–47]). The advantage of FNN is that with as few as a single hidden layer and arbitrary bounded and smooth activation functions, the system can approximate a continuous nonlinear function. The adopted model of the neuron system is represented in Figure 4.

The input data in the input layer are transferred to each neuron in the hidden layer through a linear sum operation, and the result of inputting the linear sum to the activation function is the result of the hidden layer neuron. The output of a neuron can be functionally expressed as equation (2), which is represented in Figure 4.

$$\text{out} = f(n), \quad (2)$$

where $n = \sum_{j=1}^R \omega_j x_j + b$; x_1, x_2, \dots, x_R are the input signals; $\omega_1, \omega_2, \dots, \omega_R$ are the weights of neuron; b is bias value; and $f(\cdot)$ is the network activation function.

The most common activation functions are the linear and sigmoid functions and are given according to the function $f(n) = n$ [48]. In this study, the three-layer neural network shown in Figure 3 with hyperbolic tangent neurons in the hidden layer and linear neuron in the output layer is used to simulate and approximate the dissolved oxygen. The inputs x_1, x_2, \dots, x_R are multiplied by weights $\omega_{i,j(1)}$ and summed at each hidden neuron i . Then, the summed signal $n_{i(1)} = \sum_{j=1}^R \omega_{i,j(1)} x_j + b_{i(1)}$ at a node activates a nonlinear function f_1 . The output $y = \text{DO}_t$ as a linear output node three-layered FNN is calculated according to equation (3) and generalized as shown in equation (4):

$$y = \sum_{i=1}^z \omega_{1,i(2)} \times \left(\frac{1 - e^{-\left(\sum_{j=1}^R x_j \omega_{i,j(1)} + b_{i(1)}\right)}}{1 + e^{-\left(\sum_{j=1}^R x_j \omega_{i,j(1)} + b_{i(1)}\right)}} \right) + b_{1(2)}, \quad (3)$$

$$y_k = f_o \left[\sum_{j=1}^M \omega_{jk} \times f_h \left(\sum_{i=1}^N \omega_{ij} x_i + b_j \right) + b_k \right], \quad (4)$$

where R = total number of inputs; z = hidden neurons; $\omega_{i,j(1)}$ = weight of first layer between the input j and the i th hidden neuron; $\omega_{i,j(2)}$ = weight of second layer between the i th hidden neuron and output neuron; $b_{i(1)}$ = bias weight for the i th hidden neuron; and $b_{1(2)}$ = bias weight for the output neuron.

Introduced by Rumelhart et al. [49], the backpropagation learning algorithm is the most used training algorithm for updating the weights and biases of a neural network. The network was trained on the backpropagation,

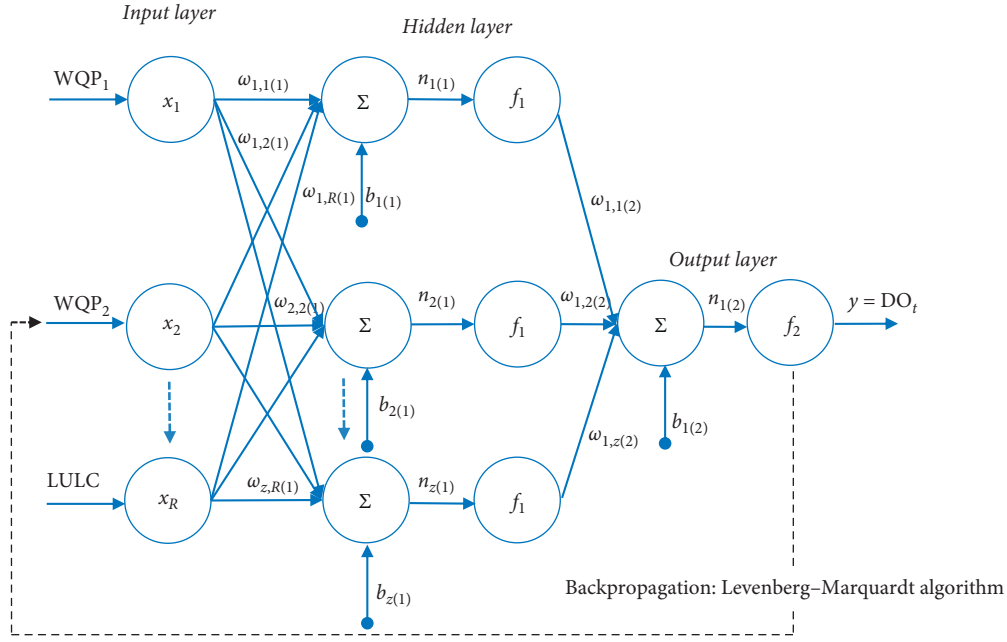


FIGURE 4: Architecture of the adopted feedforward backpropagation neural network (FNN).

which is based on a gradient scheme for weighting adjustment to reduce the error between predicted and observed data. Several variants of the backpropagation training scheme have been developed [50]; among these, the Levenberg–Marquardt algorithm is applied in this study. According to the Levenberg–Marquardt algorithm, the weights are adjusted as follows [51]:

$$E(\omega) = \frac{1}{2} \sum_{i=1}^N (y_{mi} - y_i)^2 = \frac{1}{2} (y_m - y)^T (y_m - y), \quad (5)$$

where y_i and y_{mi} are the network output and the observed value from the i th element; N is the number of training set elements; and ω is an n -element ($n = R \cdot z + 2 \cdot z + 1$) vector that contains the neural network weights and biases and is expressed as follows:

$$\omega = [\omega_{1,1(1)}, \omega_{1,2(1)}, \dots, \omega_{z,R(1)}, b_{1(1)}, b_{2(1)}, \dots, b_{z(1)}, \omega_{1,1(2)}, \omega_{1,2(2)}, \dots, \omega_{1,z(2)}, b_{1(2)}]^T. \quad (6)$$

With the Levenberg–Marquardt method, for backpropagation, the increment $\Delta\omega$, by minimization of E with respect to the weight ω parameter vector, is expressed as follows:

$$\Delta\omega = -(J^T J + \mu I)^{-1} \cdot J^T (y_m - y), \quad (7)$$

where J is the Jacobian matrix, I is the identity matrix, and μ is an adaptive factor. When the scalar μ is zero, this is just Newton's method. When μ is large, this becomes gradient descent with a small step size, and the Jacobian matrix is calculated as follows:

$$J(\omega) = \begin{bmatrix} \frac{\partial y_1}{\partial \omega_{1,1(1)}} & \dots & \frac{\partial y_1}{\partial \omega_{1,2(2)}} \\ \vdots & \vdots & \vdots \\ \frac{\partial y_N}{\partial \omega_{1,1(1)}} & \dots & \frac{\partial y_N}{\partial \omega_{1,2(2)}} \end{bmatrix}. \quad (8)$$

The input data were preprocessed through standardization within the range [0–1]. The confinement or standardization of data between limits minimizes biases and ensures all the input data receive the same attention. The data are divided into three sets, which are the training set (70%), the validation set (15%), and the test set (15%). In the determination of the ANN input data structure, different scenarios of input data were tested individually and through combinations. To avoid the selection of input data on a trial-and-error basis, cross correlation was used to determine the significant input water quality parameters.

2.6. Model Evaluation. Various methods have been used to determine the relative importance and contribution of the input variables to the model output [46, 52]. In this paper, the sensitivity analysis, based on Pearson correlation coefficient, is used to determine the influence of input variables on the dependent variable [19]. Pearson correlation coefficient is defined as the degree of correlation between the experimental and modeled values:

$$r = \frac{\sum_{i=1}^N (y_i - \bar{y})(y_{mi} - \bar{y}_m)}{\sqrt{\sum_{i=1}^N (y_i - \bar{y})^2 \sum_{i=1}^N (y_{mi} - \bar{y}_m)^2}}, \quad (9)$$

where y_i and y_{mi} denote the network output and measured value from the i th element; \bar{y} and \bar{y}_m denote network and observed averages; and N represents the number of observations.

To quantify the reliability and accuracy of the two models, FNN and MLR, the coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) were also used [53], as defined in equations (10)–(12). Additionally, the t -test method was applied to compare the

predicted value (DO_{pred}) from MLR and BPNN models, with the observed value DO_{obs} .

$$R^2 = \frac{(\sum_{i=1}^n (y_i - \bar{y})(y'_i - \bar{y}'))^2}{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (y'_i - \bar{y}')^2} \quad (10)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - y'_i)^2}{n}} \quad (11)$$

$$MAE = \frac{\sum_{i=1}^n (y'_i - y_i)}{n} \quad (12)$$

where y'_i = predicted DO by the model; y_i = target DO at time i ; and \bar{y} and \bar{y}' denote average observed and simulated DO at time i for n data point numbers.

Figure 5 presents a summary of the methodological approach in this study.

3. Results and Discussion

3.1. Correlation Analysis between Water Quality Parameters.

Table 2 presents the results of the correlational analysis of the water quality parameters. It is observed that all the water quality parameters have a negative correlation with dissolved oxygen, except for the TSS with a weak positive correlation coefficient of 0.0504. The observed correlations are discussed below.

- (a) Discharge: it is a fundamental stream property that affects most parameters from temperature to dissolved oxygen. Discharge being a function of both stream volume and speed affects DO two-fold. First, an increase in stream volume results in a decrease in the DO values since less water volume is exposed to the surface for aeration. However, an increase in stream volume also implies that a comparatively smaller surface will be exposed to the weather conditions such as temperature, and hence the stream temperature will not increase significantly leading to a relatively low DO concentration within the channel. Secondly, the speed of the streamflow significantly affects the DO concentrations in that an increase in speed results in rapid turbulence and small hydraulic jumps that facilitates the aeration process, resulting in high DO concentrations. Low flowing streams have low DO concentrations since the effects of aeration are not significantly pronounced. In the case of Nyando basin, discharge has a net negative correlation (effect) with DO (weak) meaning that an increase in stream discharge results in lower DO concentrations.
- (b) Temperature: it has a negative correlation (inverse relationship) with DO implying that an increase in temperature results in a corresponding decrease in dissolved oxygen concentrations. More dissolved oxygen is present in water with lower temperature than in warmer water. This is because the solubility of gases in liquids is an equilibrium phenomenon.

- (c) pH: the inverse correlation between DO and pH could be attributed to the fact that pH indicates the concentration of H^+ and OH^- ions in water. The H^+ ions form hydrogen bonds with water to form H_3O^+ meaning more H^+ ions result in more hydrogen bonding, thus leading to a stable structure and ultimately little free dissolved oxygen (DO).
- (d) Turbidity and total suspended solids: these are the conditions resulting from suspended solids in the water including silts and clay industrial wastes among other particles. These particles absorb heat during sunny days, thus raising the overall water temperature, which in turn lowers the dissolved oxygen levels.
- (e) Electrical conductivity: as a measure for the ability of the water to conduct electricity, EC is majorly affected by the dissolved solids that aid the transfer of electric current. A major indicator of conduction is the salinity. Salinity is important in that it affects the dissolved oxygen solubility. The higher the salinity or EC levels, the lower the dissolved oxygen concentration.
- (f) Total phosphates and total nitrates: it is observed that TP has the highest correlation with DO. There is a significant inverse correlation between phosphates and DO. Too much algae leads to a reduction of dissolved oxygen in the water since they are used up for photosynthesis. Phosphates are present in fertilizers and normally enter the water bodies through agricultural runoff or as sewage discharge. An increase in the amount of phosphates present in water results in a corresponding decrease in the amount of DO. The presence of nitrates in natural waters has the same effects as those of phosphates.

3.2. Correlation between LULC and Dissolved Oxygen.

Table 3 presents the results of the correlation analysis between LULC classes and average concentration of the DO values, and it is observed that the LULC within the watershed has significant impact on the DO water quality of the streams. The stream water quality fluctuates with changes in the LULC parameters within the contributing watershed. Notably, changes in the land cover and land management practices have been regarded as the key influencing factors behind the alteration of the hydrological system, which leads to change in runoff as well as water quality. From the results, forest cover, light vegetation, and agricultural lands have significantly stronger correlation with water quality (DO), positively or negatively. Negative correlation implies, for example, as agricultural land increases, the forest cover decreases strongly with a correlation factor of -0.97552 . Similarly, positive correlation means that as forest cover increases, concentration of DO increases in the rivers by a correlation factor of $+0.83559$.

3.2.1. Forest and Light Vegetation Cover. With up to 83.56% correlation between DO and forest cover, it is inferred that increase in vegetation cover results in increased DO

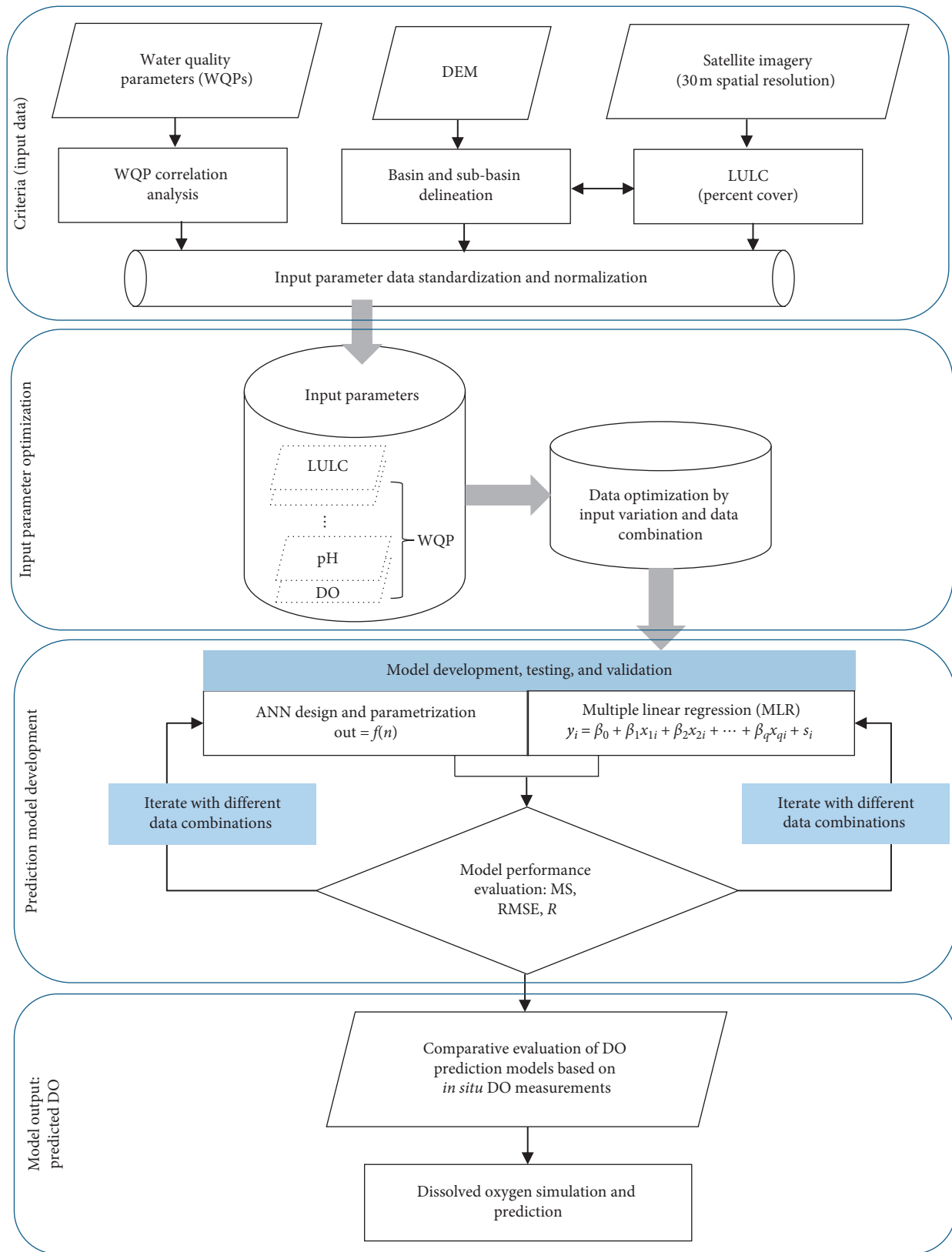


FIGURE 5: Schematic workflow of the approach for DO prediction using ANN and MLR models.

concentrations within the basin. Vegetation cover redistributes rainfall into three components (i.e., canopy interception, streamflow, and throughfall), thereby weakening

rainfall, limiting runoff occurrence, and controlling soil erosion. In general, runoff and sediment yields decrease with coverage rates as a linear or exponential function. With

TABLE 2: Correlational analysis between DO and the WQ parameters.

	Discharge	Temp	pH	Turbidity	EC	TP	TN	TSS	DO
Discharge	1								
Temp	0.271748	1							
pH	-0.22542	0.069421	1						
Turbidity	0.53972	0.287929	-0.28211	1					
EC	0.190513	0.282863	0.352155	-0.14834	1				
TP	0.133385	0.194775	0.114777	0.185624	0.237535	1			
TN	0.034775	-0.00622	-0.03917	0.016368	0.283277	0.231251	1		
TSS	0.689712	0.260862	-0.34453	0.818282	-0.02368	0.119882	-0.03906	1	
DO	-0.13954	-0.20597	-0.24034	-0.09428	-0.27706	-0.4543	-0.34444	0.05042	1

TABLE 3: Correlation between LULC and DO in the Nyando River basin.

	Forest	Wetlands	Light veg	Agric land	DO (mg/l)
Forest	1				
Wetlands	0.495085	1			
Light veg	-0.16694	-0.45426	1		
Agric land	-0.97552	-0.43501	-0.04913	1	
DO (mg/l)	0.835591	-0.00072	0.277317	-0.89189	1

increase in vegetation cover, runoff and soil erosion are controlled and other water quality parameters such as TS, EC, and nutrient load (phosphates and nitrates) discharged into streams are reduced.

3.2.2. Agricultural Land and Activities. The strong negative correlation between agricultural land cover and dissolved oxygen concentration observed in the Nyando basin streams implies that an increase in the percentage of land under agriculture leads to a significant decrease in the DO concentrations. This is majorly due to the nonpoint source pollution due to the runoff from agricultural lands. Fertilizers, rich in phosphates, nitrates, and other nutrients, and other farm chemicals such as pesticides and herbicides are washed off by runoff into receiving water bodies within the watershed. Increased nutrients' concentration into streams has its adverse effects on the water quality in the streams as presented in the WQP correlational analysis in Table 2.

3.3. Dissolved Oxygen Prediction Results Using the MLR Model. The results for the prediction of DO using the water quality parameters in the basin using MLR are presented in Table 4, with the results showing that the combination of EC, TP, TN, and pH water quality variables being the four best performing water quality parameters in the prediction of dissolved oxygen concentrations. It is observed that the optimal prediction of DO using MLR was obtained by incorporating all the input parameters, resulting in R of 61.99%, which is not significantly higher than 57.03% when only four parameters are used.

The best results for the prediction of DO using MLR based on the water quality parameters and LULC are, respectively, represented in equations (13) and (14), with the water quality predicting the DO concentration with $R=0.6199$, while LULC was a more accurate predictor with $R=0.9917$ (Table 4). The total degrees of freedom df_{total} for

the regression models numbered 1–4 in Table 4 were 56, with the residual df_{res} being equivalent to $df_{total} - df_{reg} - 1$, and regression df_{reg} is equivalent to the number of model input parameters for each estimation. For the LULC-based regression (equation (14)), $df_{total} = 16$, $df_{reg} = 4$, and $df_{res} = 11$ for the eight sub-basins including the main basin.

$$\begin{aligned} DO = & 12.22 - 0.023\text{Discharge} - 0.032\text{Temp} - 0.342\text{pH} \\ & - 0.002\text{Turb} - 0.0003\text{EC} - 0.847\text{TP} - 0.244\text{TN} \\ & + 0.002\text{TSS}, \end{aligned} \quad (13)$$

$$\begin{aligned} DO = & 2.6192 + 0.0909\text{Forest} - 0.2372\text{Wetlands} \\ & + 0.0909\text{Light Veg} + 0.0449\text{Agric Land}. \end{aligned} \quad (14)$$

The results in Table 4 show that the combination of temperature, pH, electrical conductivity, total nitrates, and total phosphates predicted DO with an accuracy of 57.03%, while the discharge, turbidity, and TSS have insignificant influence on the prediction of DO with only 4% contribution. For LULC, agricultural land and forest land cover had the highest impacts on DO prediction.

Figure 6 shows the simulation of DO concentrations based on the different input parameters in comparison to the actual observed DO for all the sampling stations used in the study. For all the input stations, the simulation trend is observed to follow the observed trend. However, for some stations, the magnitudes of the observed DO concentrations tend to be higher than that for the simulated, even when all the available water quality variables are used. This could imply that not all the water quality parameters that are required for the simulation and prediction of DO were available, hence the observed trend with best $R=0.6199$. From the moving average trend line in Figure 6, there is an indication that the general trend in the DO concentration within the basin is decreasing.

TABLE 4: MLR performance and model equations for water quality parameters and LULC.

Input model parameters	R	Adjusted R^2	MSE	RMSE	MAE	MLR model equation
1 EC, TP, TN	0.53000	0.240199	0.61105	0.78169	0.58821	DO = 8.9268 - 0.0013EC - 0.9887TP - 0.2292TN
2 EC, TP, TN, pH	0.55835	0.258816	0.58483	0.76474	0.58188	DO = 11.685 - 0.0006EC - 0.9578TP - 0.2609TN - 0.3633pH
3 EC, TP, TN, pH, temperature	0.57025	0.259036	0.57341	0.75724	0.58235	DO = 12.71 - 0.0002EC - 0.9056TP - 0.2776TN - 0.3776pH - 0.0449Temp
4 All WQ parameters	0.61999	0.281797	0.52310	0.72326	0.53699	DO = 12.22 - 0.023Discharge - 0.032Temp - 0.342pH - 0.002Turb - 0.0003EC - 0.847TP - 0.244TN + 0.002TSS
LULC categories: forest; wetlands; light veg; agric land	0.9917	0.96147	0.00085	0.02929	0.02266	DO = 2.6192 + 0.0909Forest - 0.2372Wetlands + 0.0909Light Veg + 0.0449Agric Land

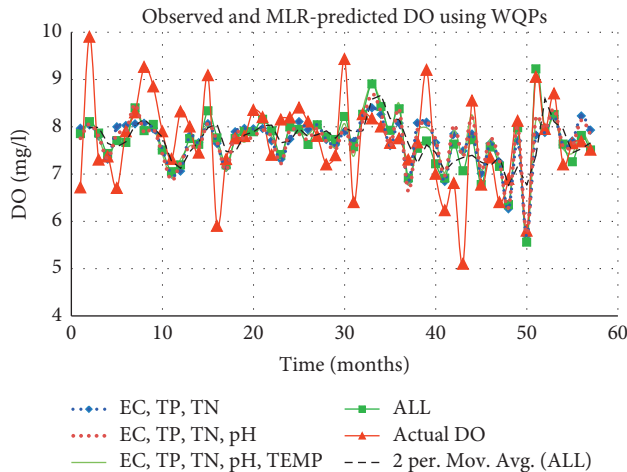


FIGURE 6: Variations of the observed and MLR-predicted DO concentrations (ALL implies all the water quality parameters as inputs).

The stream dissolved oxygen concentration is a complex phenomenon affected by factors that are ever changing, some of which may not entirely be captured by traditional laboratory measurement techniques. The DO concentration in streams is never constant even under the most stable atmospheric conditions such as temperature, rainfall, and wind velocity among others. Increasing the parameters under consideration improves the model performance by bringing into view other factors that may not have been considered before.

By using the LULC percentages in the prediction of DO using MLR, the results in Table 4 indicate that LULC is a perfect predictor of the DO concentrations within the Nyando River, with $R^2 = 0.9915$ and negligible RMSE, MSE, and MAE. For the eight sub-basins, the comparison between the predicted and actual DO concentrations is presented in Figure 7. While there is an indication of marginal DO concentration decreasing with time as depicted in Figure 6, the overall trend in Figure 7 shows the DO concentration in the respective sub-basins is cumulatively increasing from upstream to downstream. The observed increase in the DO concentrations is attributed to lower chemical flows into the streams in the upstream, and this accumulates as there are intensified agricultural, industrial, and human settlements downstream. In this region, there is minimal control of the effluent into the river, and this results in observed trend, indicating that the critical areas are situated in the downstream sections of the basin.

By combining LULC percent and the average of the sub-basin water quality parameters (EC, TP, and TN), DO was predicted with an accuracy of $R = 0.9924$, which is equivalent to predicting the DO in the river using LULC only.

The LULC affects more parameters contributing to stream DO concentrations within the basin. Land practices have direct impacts on the velocity and amount of runoff flowing into receiving streams, nutrients, total solids, and pH of the streams, and these factors have direct impacts on the DO concentrations and hence the improved model performance. Compared to the WQ parameters, it is

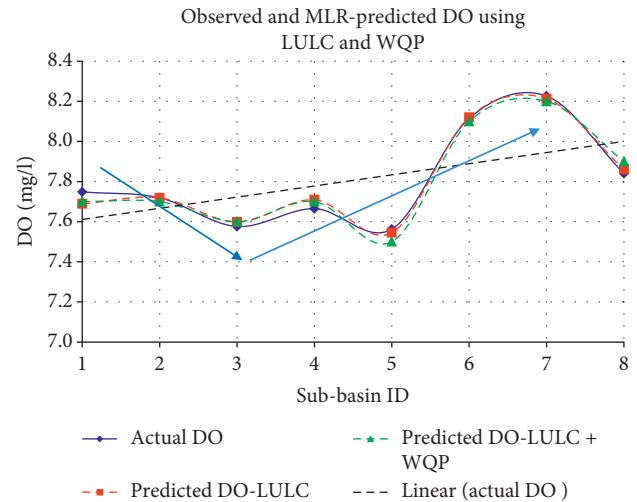


FIGURE 7: Observed and MLR-predicted DO concentrations using LULC in each sub-basin.

conclusive that LULC is a better predictor of DO in river basins with inadequate and unreliable water quality monitoring networks. The combined results of the LULC and three WQ parameters show the same trend as the actual and LULC predicted DO.

3.4. Prediction of Dissolved Oxygen Using the FNN Model

3.4.1. Optimum Neuron Determination. By varying the input water quality variables through using trial-and-error and also varying the neurons between 10 and 50, 25 neurons were found to be optimum as depicted in Figure 8. When the neurons are too few, e.g., 1–10 neurons, the network lacks the capacity to sufficiently learn the underlying data patterns and to detect signals in the dataset, and this resulted in underfitting with higher RMS error values. When the neurons are increased to 50, the network performance does not significantly increase. As depicted in Figure 8, the optimal prediction based on 3-data combination input is at 25 neurons. Using the Levenberg–Marquardt training function, the performance of the network's weights and biases in terms of mean square error was set at 8 epochs, with a learning rate of 0.1 (Figure 9).

3.4.2. DO Prediction Using FNN with Varying Water Quality Parameters. To model the influence and correlation of the different WQ parameters on the concentration of DO within the river, different combinations of water quality parameters were used as inputs in the FNN prediction model. By first using EC, TP, and TN, the DO was predicted with R values of 0.7203 for training and 0.8097 for validation, which further increased to 0.9132 for the testing datasets and an overall average R value of 0.7660 (Figure 10(a)). With the inclusion of pH together with the three parameters, the DO prediction improved by about 18% according to the correlation coefficient; however, R reduced to 0.6013 after validation and increased to 0.9395 during testing (Figure 10(b)). Using the four WQPs, the overall average R value of 0.8141 was

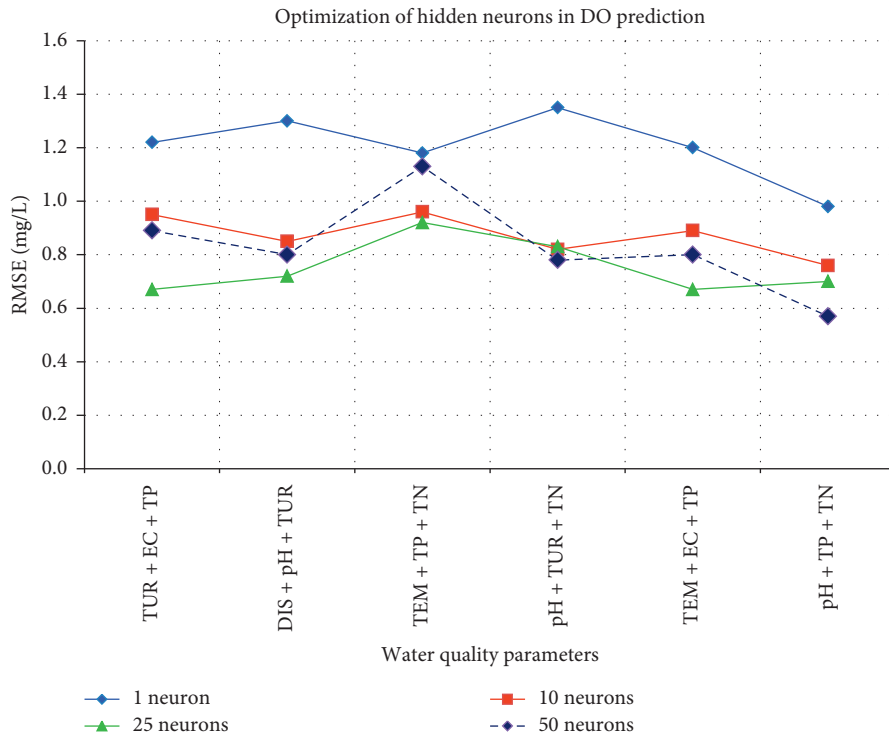


FIGURE 8: Performance evaluation for the selection of hidden layer for DO prediction using FNN.

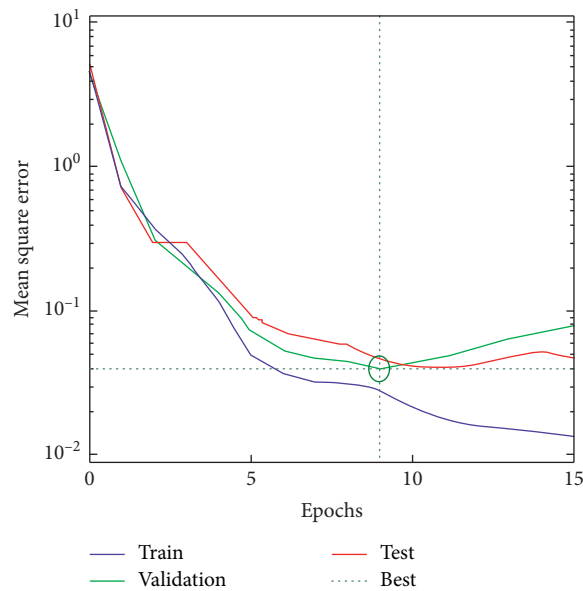


FIGURE 9: Training of the RNN model.

obtained, which increased approximately by 5% in the DO prediction accuracy compared to the three WQPs.

By successively increasing the number of the input WQ parameters in predicting the DO concentration, the results progressively improved as shown in Figure 10, with a maximum prediction accuracy of $R = 0.8546$ being obtained by inputting all the eight water quality variables (Figure 10(d)). A graphical representation of the correlations between the predictions of DO using ANN and the

observed DO concentrations is given in Figure 11. It is observed that as the number of input water quality parameters increases, the accuracy of prediction of the ANN also increases. Comparing the model results in Figures 11, 6, and 7, it is observed that the optimized FNN simulation model follows the pattern of the observed data, with very few outliers when the input variables are few, confirming the FNN models yield better results with fewer input variables.

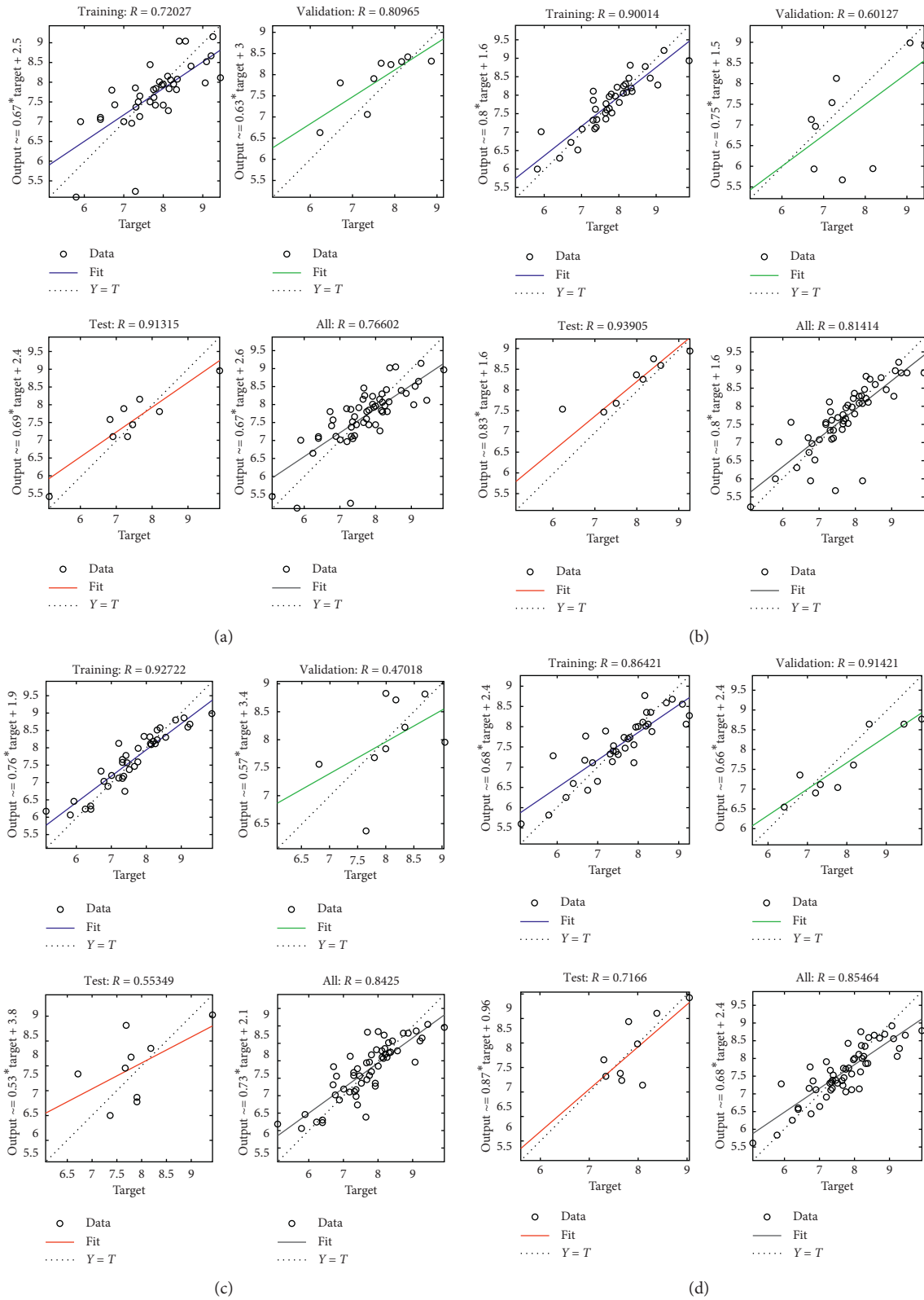


FIGURE 10: FNN-based predictions of DO with (a) EC, TP, and TN; (b) EC, TP, TN, and pH; (c) EC, TP, TN, pH, and temperature; and (d) eight (8) input WQ parameters.

In Figure 11, the average trend line also shows that there is a general decrease in the concentration of DO within the basin with time. The same phenomenon is observed in

Figure 6, where the moving average is generally decreasing. This observed decrease during the study period is however within the acceptable range but with negative magnitude.

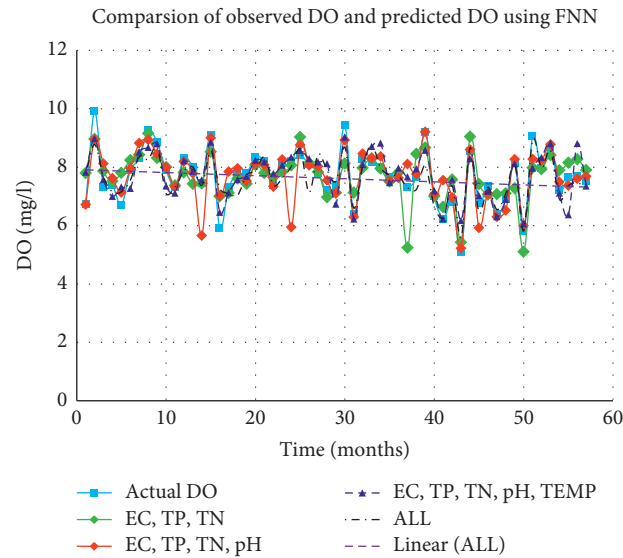


FIGURE 11: Observed DO and FNN-predicted DO for all the input samples within the basin (the ALL trend is for LULC and water quality parameters and its linear estimation is the Linear (ALL)).

Causes thereof need to be further investigated, through continuous monitoring and predictions of water quality within the basin.

By introducing the LULC percentages in the prediction of the DO within the basin, the results in Figure 12(a) show that LULC has a higher predictive ability of dissolved oxygen with as much as 99.5% accuracy. Similarly, the combination of LULC and WQPs yields equally satisfactory correlation results with $R^2 = 0.997$ (Figure 12(b)), implying that in the absence of adequate water quality parameters, LULC can sufficiently be used to predict and estimate the concentrations of DO within a basin with low water quality monitoring networks. For the prediction of DO using LULC and combined LULC with the WQPs, the MSEs were, respectively, as low as 0.0665 mgL^{-1} and 0.0005 mgL^{-1} , while for the same combination, the RMSEs were, respectively, at 0.2578 mgL^{-1} and 0.0218 mgL^{-1} . Also the MAE for the two input models is, respectively, observed as 0.1847 mgL^{-1} and 0.1878 mgL^{-1} . The FNN results in Figure 12(a) are same as using the MLR model in DO prediction presented in Figure 5. This means that in the absence of water quality parameters, LULC can satisfactorily predict DO concentrations. However, with scarce WQPs, it is preferable to use the ANN model. For FNN, it can also be deduced that few input parameters, say five, produce considerably acceptable results, provided the parameters used have a strong correlation with DO. The difference is more conspicuous in the MSE, which displays a value of 0.0665 mgL^{-1} for when only LULC is used, versus 0.0005 mgL^{-1} for when both LULC and WQ parameters are used.

3.5. Further Discussion. The statistical MLR, by relating the dependent variable to independent variables, has the weakness that the transformations include *a priori* assumptions about the type and consistency of the relation

between two parameters which may not be met completely [54]. This could contribute to its inferior performance in the prediction of DO concentrations, as compared to the FNN model. For both FNN and MLR models, LULC presented superior results as compared to the water quality parameters. This means that the use of LULC may be more significant indicator in DO prediction and as such watershed models may rely on the effects of LULC on water quality.

Despite a number of data-driven models being proposed for DO prediction, little attention has been paid to developing systematic ways for the selection of appropriate model inputs [55]. Sarkar and Pandey [23], for example, combined datasets from three different monitoring stations and used them as input to the ANN model, without any feature selection strategy. Shi et al. [56], on the other hand, input seven surface water quality variables, without any *a priori* variable analysis. For optimized neural network modeling, the inclusion of all available variables may contain redundant input parameters, which can decrease the performance of the model. To take into consideration the linear dependencies, simple evaluations of inputs through trial-and-error need to assess all possible groups of inputs by building a number of predictive models, which can be inefficient when dealing with large input water quality variables.

Comparatively, satisfactory DO simulation and prediction results have been obtained in this study. For example, it is observed that as the input parameters increase, the accuracy of the outputs of the models increases at all the stages of training, testing, and validation. The trend depicts saturation in the prediction accuracy as the correlated water quality variables are introduced into the model. The lower performance of the water quality parameters in the prediction DO may be attributed to the nonhomogenous nature of the water quality variables and also due to the fact the input parameters in this study may not include all the relevant variables suited for DO prediction, and the models also required high spatially distributed and long-temporal

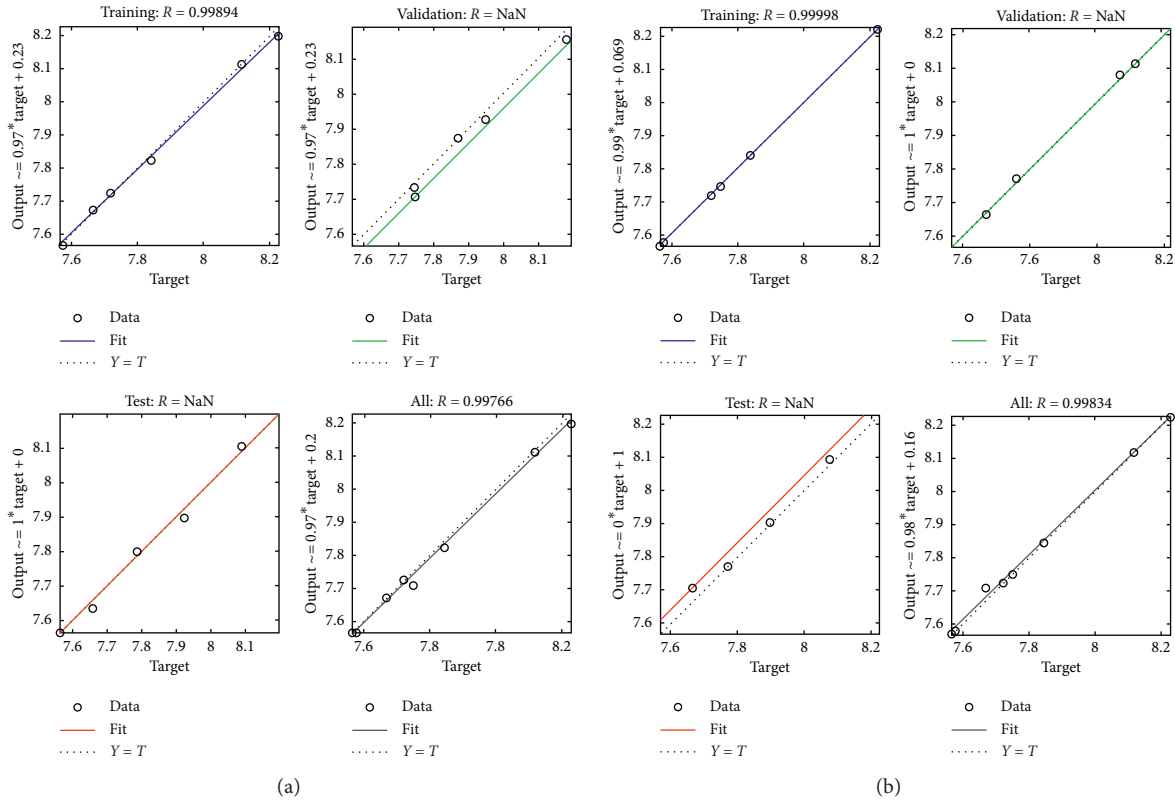


FIGURE 12: FNN model prediction of DO concentrations using (a) LULC and (b) combination of LULC and the eight (8) WQ parameters.

observation water quality data. By introducing LULC, the current study significantly improves on the DO prediction as the results depict perfect simulation and prediction. This however may be improved and investigated further by comparing different temporal seasons for LULC data. For practical applications, such models need to be calibrated or fine-tuned with *in situ* observations [57, 58].

Compared to previous studies in [1, 18, 21–23, 35, 36], it is recognized that only the study by Kanda et al. [22] was in the same geographic region and used nearly similar input model parameters in the prediction of DO. The results obtained in the current study are however superior to those obtained by Kanda et al. [22] in terms of the accuracy of DO prediction and because of the incorporation of LULC variability as proposed by Kalin and Isik [35] and Ahearn et al. [36]. This implies that while most studies rely on BOD and COD for prediction of DO, in poorly monitored river basins, LULC and other water quality parameters as used in this study can be incorporated into FNN models for the accurate prediction of DO.

The proposed FNN model shows efficiency in forecasting the dissolved oxygen profiles in eutrophic river water bodies as the Nyando River basin. The FNN model can thus be used for forecasting, so as to capture long-term trends observed for the tedious water quality variables such as dissolved oxygen [10, 33]. However, because water quality predictions can be easily affected with high uncertainty and specific phenomena, such as climatological and ecoregional conditions, the predictions should be applied with care as they can exhibit certain deviations. In the implementation of the

model, the selection of the optimum number of hidden neurons to be used in the ANN model is a significant step as it either causes improvement in performance or a decrease in the same, that is, underfitting or overfitting.

4. Conclusions

Modeling water quality variables is significant in the analysis of aquatic systems. However, the chemical, physical, and biological components of aquatic ecosystems vary and are complex and nonlinear in the relationship. The study results showed that the application of feedforward backpropagation neural networks (FNNs) is an effective approach in the identification and modeling of nonlinear interacting water quality parameters for the prediction of dissolved oxygen in scarcely monitored basins, as compared with the statistical multiple linear regression (MLR).

Correlational analysis for the optimization of the input parameters and minimization of the redundancies in the input water quality parameters into the prediction models showed that the parameters that have stronger relationships with DO are the most significant in its prediction for both the MLR and FNN models. The study results show that in the prediction of DO using water quality parameters, optimal results were obtained by combining temperature, electric conductivity, total phosphorus, pH, and total nitrates as the predictor variables for both FNN and MLR models, with correlation coefficients $R = 0.8425$ and $R = 0.5703$, respectively. By including all the available water quality parameters, both models improved marginally in the DO

concentration predictions with $R=0.8564$ using FNN and $R=0.6200$ using the statistical MLR. The results showed that FNN outperformed MLR by 24% using water quality parameters only.

By using LULC to predict the DO concentration in the river, the modeling results show that both the models performed at 99% prediction accuracy. A combination of LULC and water quality only showed insignificant improvement in the DO prediction. The results show that the proposed optimized FNN is an efficient alternative for the modeling of the variability of water quality parameters in basins which are scarcely monitored. While it is agreed that unmonitored watersheds are faced with the problem of inadequate data, for the modeling systems to work, it is important to improve on the development of the models using long-term data as this will improve the reliability and accuracy of the model output. With more data over long-term temporal observations, deep learning neural networks can then be employed in the development of the artificial intelligence model.

Data Availability

The data used to support the findings of this study are included within the article.

Disclosure

The authors carried out the research while at the Department of Civil and Structural Engineering, Moi University, Kenya.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors would like to acknowledge the Lake Victoria South Water Service Board (LVSWSB), Kisumu (Kenya), for providing the water quality data used in this study.

References

- [1] M. Bonansea, C. Ledesma, and M. C. Rodriguez, "Assessing the impact of land use and land cover on water quality in the watershed of a reservoir," *Applied Ecology and Environmental Research*, vol. 14, no. 2, pp. 441–456, 2016.
- [2] M. E. Meding and L. J. Jackson, "Biotic, chemical, and morphometric factors contributing to winter anoxia in prairie lakes," *Limnology and Oceanography*, vol. 48, no. 4, pp. 1633–1642, 2003.
- [3] R. D. Robarts, M. J. Waiser, M. T. Arts, and M. S. Evans, "Seasonal and diel changes of dissolved oxygen in a hypertrophic prairie lake," *Lakes and Reservoirs: Research and Management*, vol. 10, no. 3, pp. 167–177, 2005.
- [4] P. Branco, J. M. Santos, S. Amaral, F. Romão, A. N. Pinheiro, and M. T. Ferreira, "Potamodromous fish movements under multiple stressors: connectivity reduction and oxygen depletion," *Science of the Total Environment*, vol. 572, pp. 520–525, 2016.
- [5] Z. Xiao, L. Peng, Y. Chen, H. Liu, J. Wang, and Y. Nie, "The dissolved oxygen prediction method based on neural network," *Complexity*, vol. 2017, pp. 1–6, 2017.
- [6] W.-B. Chen and W.-C. Liu, "Artificial neural network modeling of dissolved oxygen in reservoir," *Environmental Monitoring and Assessment*, vol. 186, no. 2, pp. 1203–1217, 2014.
- [7] S. R. Poulson and A. B. Sullivan, "Assessment of diel chemical and isotopic techniques to investigate biogeochemical cycles in the upper Klamath River, Oregon, USA," *Chemical Geology*, vol. 269, no. 3–4, pp. 3–11, 2010.
- [8] C. Lihua, M. Shengquan, and L. I. Li, "A model to evaluate do of river based on artificial neural network and style book," *Journal of Hainan Normal University (Natural Science)*, vol. 21, pp. 372–376, 2008.
- [9] J. Cho and S. R. Ha, "Parameter optimization of the QUAL2 K model for a multiple-reach river using an influence coefficient algorithm," *Science of the Total Environment*, vol. 408, no. 8, pp. 1985–1991, 2010.
- [10] K. P. Singh, A. Basant, A. Malik, and G. Jain, "Artificial neural network modeling of the river water quality—a case study," *Ecological Modelling*, vol. 220, no. 6, pp. 888–895, 2009.
- [11] X. Ji, X. Shang, R. A. Dahlgren, and M. Zhang, "Prediction of dissolved oxygen concentration in hypoxic river systems using support vector machine: a case study of Wen-Rui Tang River, China," *Environmental Science and Pollution Research*, vol. 24, pp. 16061–16076, 2017.
- [12] S. Zhu and S. Heddam, "Prediction of dissolved oxygen in urban rivers at the three gorges Reservoir, China: extreme learning machines (ELM) versus artificial neural network (ANN)," *Water Quality Research Journal*, vol. 55, no. 1, pp. 106–118, 2020.
- [13] N. Basant, S. Gupta, A. Malik, and K. P. Singh, "Linear and nonlinear modeling for simultaneous prediction of dissolved oxygen and biochemical oxygen demand of the surface water—a case study," *Chemometrics and Intelligent Laboratory Systems*, vol. 104, no. 2, pp. 172–180, 2010.
- [14] V. Ranković, J. Radulović, I. Radojević, A. Ostojić, and L. Čomić, "Neural network modeling of dissolved oxygen in the Gruža reservoir, Serbia," *Ecological Modelling*, vol. 221, no. 8, pp. 1239–1244, 2010.
- [15] Y. O. Ouma, J. Waga, M. Okech, O. Lavisa, and D. Mbutia, "Estimation of reservoir bio-optical water quality parameters using smartphone sensor apps and Landsat ETM+: review and comparative experimental results," *Journal of Sensors*, vol. 2018, Article ID 3490757, 32 pages, 2018.
- [16] K.-W. Chau, "A review on integration of artificial intelligence into water quality modelling," *Marine Pollution Bulletin*, vol. 52, no. 7, pp. 726–733, 2006.
- [17] Z. He, X. Wen, H. Liu, and J. Du, "A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region," *Journal of Hydrology*, vol. 509, pp. 379–386, 2014.
- [18] A. A. M. Ahmed, "Prediction of dissolved oxygen in Surma River by biochemical oxygen demand and chemical oxygen demand using the artificial neural networks (ANNs)," *Journal of King Saud University-Engineering Sciences*, vol. 29, no. 2, pp. 151–158, 2017.
- [19] E. Dogan, B. Sengorur, and R. Koklu, "Modeling biological oxygen demand of the Melen river in Turkey using an artificial neural network technique," *Journal of Environmental Management*, vol. 90, no. 2, pp. 1229–1235, 2009.

- [20] S. Palani, S.-Y. Liong, and P. Tklich, "An ANN application for water quality forecasting," *Marine Pollution Bulletin*, vol. 56, no. 9, pp. 1586–1597, 2008.
- [21] S. I. Abba, S. J. Hadi, and J. Abdullahi, "River water modelling prediction using multi-linear regression, artificial neural network, and adaptive neuro-fuzzy inference system techniques," *Procedia Computer Science*, vol. 120, pp. 75–82, 2017.
- [22] E. K. Kanda, E. C. Kipkorir, and J. R. Kosgei, "Dissolved oxygen modelling using artificial neural network: a case of river Nzoia, Lake Victoria basin, Kenya," *Journal of Water Security*, vol. 2, no. 1, pp. 1–7, 2016.
- [23] A. Sarkar and P. Pandey, "river water quality modelling using artificial neural network technique," *Aquatic Procedia*, vol. 4, pp. 1070–1077, 2015.
- [24] S. Huo, Z. He, J. Su, B. Xi, and C. Zhu, "Using artificial neural network models for eutrophication prediction," *Procedia Environmental Sciences*, vol. 18, pp. 310–316, 2013.
- [25] M. Heydari, E. Olyaie, H. Mohebzadeh, and Ö. Kisi, "Development of a neural network technique for prediction of water quality parameters in the Delaware river, Pennsylvania," *Middle-East Journal of Scientific Research*, vol. 13, no. 10, pp. 1367–1376, 2013.
- [26] X. Wen, J. Fang, M. Diao, and C. Zhang, "Artificial neural network modeling of dissolved oxygen in the Heihe River, Northwestern China," *Environmental Monitoring and Assessment*, vol. 185, no. 5, pp. 4361–4371, 2013.
- [27] D. Antanasijević, V. Pocajt, A. Perić-Grujić, and M. Ristić, "Modelling of dissolved oxygen in the Danube River using artificial neural networks and Monte Carlo Simulation uncertainty analysis," *Journal of Hydrology*, vol. 519, pp. 1895–1907, 2014.
- [28] K. Sinha and P. D. Saha, "Assessment of water quality index using cluster analysis and artificial neural network modeling: a case study of the Hooghly River basin, West Bengal, India," *Desalination and Water Treatment*, vol. 54, no. 1, pp. 28–36, 2015.
- [29] P. H. Dimberg and A. C. Bryhn, "Predicting total nitrogen, total phosphorus, total organic carbon, dissolved oxygen and iron in deep waters of Swedish lakes," *Environmental Modelling & Assessment*, vol. 20, no. 5, pp. 411–423, 2015.
- [30] S. Soyupak, F. Karaer, H. Gürbüz, E. Kivrak, E. Sentürk, and A. Yazici, "A neural network-based approach for calculating dissolved oxygen profiles in reservoirs," *Neural Computing & Applications*, vol. 12, no. 3-4, pp. 166–172, 2003.
- [31] P. Chaves and T. Kojiri, "Conceptual fuzzy neural network model for water quality simulation," *Hydrological Processes*, vol. 21, no. 5, pp. 634–646, 2007.
- [32] J.-T. Kuo, M.-H. Hsieh, W.-S. She, and N. Shed, "Using artificial neural network for reservoir eutrophication prediction," *Ecological Modelling*, vol. 200, no. 1-2, pp. 171–177, 2007.
- [33] Z. Ying, N. Jun, C. Fuyi, and G. Liang, "Water quality forecast through application of BP neural network at Yuquiao reservoir," *Journal of Zhejiang University-SCIENCE A*, vol. 8, pp. 1482–1487, 2007.
- [34] P. Basnyat, L. D. Teeter, B. G. Lockaby, and K. M. Flynn, "Research: land use characteristics and water quality: a methodology for valuing of forested buffers," *Environmental Management*, vol. 26, no. 2, pp. 153–161, 2000.
- [35] L. Kalin and S. Isik, "Prediction of water quality parameters using an artificial neural networks model," in *Proceedings of the World Environmental and Water Resources Congress 2010*, American Society of Civil Engineers, Providence, RI, USA, pp. 3145–3153, May 2010.
- [36] D. S. Ahearn, R. W. Sheibley, R. A. Dahlgren, M. Anderson, J. Johnson, and K. W. Tate, "Land use and land cover influence on water quality in the last free-flowing river draining the western Sierra Nevada, California," *Journal of Hydrology*, vol. 313, no. 3-4, pp. 234–247, 2005.
- [37] V. Z. Antonopoulos and S. K. Gianniou, "Simulation of water temperature and dissolved oxygen distribution in Lake Vegoritis, Greece," *Ecological Modelling*, vol. 160, no. 1-2, pp. 39–53, 2003.
- [38] L. O. Olang, P. Kundu, T. Bauer, and J. Fürst, "Analysis of spatio-temporal land cover changes for hydrological impact assessment within the Nyando River Basin of Kenya," *Journal of Environmental Monitoring and Assessment*, vol. 179, no. 1-4, pp. 389–401, 2011.
- [39] M. Razi and K. Athappilly, "A comparative predictive analysis of neural networks (NNs), nonlinear regression and classification and regression tree (CART) models," *Expert Systems with Applications*, vol. 29, no. 1, pp. 65–74, 2005.
- [40] A. Csábrági, S. Molnár, P. Tanos, and J. Kovács, "Application of artificial neural networks to the forecasting of dissolved oxygen content in the Hungarian section of the river Danube," *Ecological Engineering*, vol. 100, pp. 63–72, 2017.
- [41] S. K. Golfinopoulos and G. B. Arhonditsis, "Multiple regression models: a methodology for evaluating trihalomethane concentrations in drinking water from raw water characteristics," *Chemosphere*, vol. 47, no. 9, pp. 1007–1018, 2002.
- [42] S. Landou and B. S. Everitt, *A Handbook of Statistical Analyses using SPSS*, Chapman & Hall/CRC, Boca Raton, FL, USA, 2004.
- [43] C. W. Dawson and R. L. Wilby, "Hydrological modelling using artificial neural networks," *Progress in Physical Geography*, vol. 25, no. 1, pp. 80–108, 2001.
- [44] N. J. De Vos and T. H. M. Rientjes, "Constraints of artificial neural networks for rainfall-runoff modelling: trade-offs in hydrological state representation and model evaluation," *Hydrology and Earth System Sciences*, vol. 9, pp. 111–126, 2005.
- [45] S. M. Sumi, M. F. Zaman, and H. Hirose, "A rainfall forecasting method using machine learning models and its application to the Fukuoka city case," *International Journal of Applied Mathematics and Computer Science*, vol. 22, no. 4, pp. 841–854, 2012.
- [46] S. Singh, S. Jain, and A. Bárdossy, "Training of artificial neural networks using information-rich data," *Hydrology*, vol. 1, no. 1, pp. 40–62, 2014.
- [47] T.-W. Kim and J. B. Valdés, "Nonlinear model for drought forecasting based on a conjunction of wavelet transforms and neural networks," *Journal of Hydrologic Engineering*, vol. 8, no. 6, pp. 319–328, 2003.
- [48] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Prentice-Hall, Upper Saddle River, NJ, USA, 1999.
- [49] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.
- [50] M. T. Hagan and M. B. Menhaj, "Training feedforward networks with the Marquardt algorithm," *IEEE Transactions on Neural Networks*, vol. 5, no. 6, pp. 989–993, 1994.
- [51] J. J. Moré, "The Levenberg-Marquardt Algorithm: implementation and theory," in *Numerical Analysis*, pp. 105–116, Springer, New York, NY, USA, 1978.
- [52] R. J. May, H. R. Maier, G. C. Dandy, and T. M. Fernando, "Non-linear variable selection for artificial neural networks

- using partial mutual information,” *Environmental Modeling and Software*, vol. 23, no. 10-11, pp. 1312–1326, 2008.
- [53] S. Huang, B. Ming, Q. Huang, G. Leng, and B. Hou, “A case study on a combination NDVI forecasting model based on the entropy weight method,” *Water Resources Management*, vol. 31, no. 11, pp. 3667–3681, 2017.
- [54] T. Brey, A. Jarre-Teichmann, and O. Borlich, “Artificial neural network versus multiple linear regression: predicting P/B ratios from empirical data,” *Marine Ecology Progress Series*, vol. 140, pp. 251–256, 1996.
- [55] J. Wu, Z. Li, L. Zhu, G. Li, B. Niu, and F. Peng, “Optimized BP neural network for Dissolved Oxygen prediction,” *IFAC-PapersOnLine*, vol. 51, no. 17, pp. 596–601, 2018.
- [56] B. Shi, P. Wang, J. Jiang, and R. Liu, “Applying high-frequency surrogate measurements and a wavelet-ANN model to provide early warnings of rapid surface water quality anomalies,” *Science of the Total Environment*, vol. 610-611, pp. 1390–1399, 2018.
- [57] N. Fohrer, S. Haverkamp, K. Eckhardt, and H.-G. Frede, “Hydrologic response to land use changes on the catchment scale,” *Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere*, vol. 26, no. 7-8, pp. 577–582, 2001.
- [58] M. Luzio, R. Srinivasan, and J. G. Arnold, “Integration of watershed tools and SWAT model into BASINS,” *Journal of the American Water Resources Association*, vol. 38, no. 4, pp. 1127–1141, 2002.