Delta University Scientific Journal
Journal home page: https://dusj.journals.ekb.eg

Developing Water Quality Indices Utilizing Artificial Neural Networks: A Case Study of the "Gharbia" Main Drain in Egypt

Yasmina W. Elsirwi[1], Mahmoud A. Elsheiwy[2], Ehab A. Elsayed[2, 3], Alaa A. El-Razek[2]

[1] *Master of Engineering Science Candidate, Faculty of Engineering, Delta University for Science and Technology, Gamasa, Egypt, P.O. BOX 11152*

[2] *Civil Engineering Department, Faculty of Engineering, Delta University for Science and Technology, Gamasa, Egypt, P.O. BOX 11152*
[3] *Drainage Research Institute, National Water Research Center*

Correspondence: Yasmina W. Elsirwi, Gamasa, Egypt, P.O. Box 11152, Tel+201011110095; Email: yasminawahba7@gmail.com

ABSTRACT

Egypt faces significant challenges in managing its limited water resources. Agricultural drainage water has become a crucial unconventional water resource that can be reused for irrigation. This study focuses on the Gharbia Main Drain, a crucial water source that provides 1.9 billion cubic meters of water annually. However, it is heavily polluted by agricultural runoff, domestic wastewater, and industrial discharges. To address this, the study aims to develop accurate Water Quality Indices (WQIs) for assessing pollution levels in the drain using Artificial Neural Networks (ANNs). This study involves analyzing nineteen water quality parameters covering key biological, industrial, and agricultural pollutants. The Canadian Water Quality Index (CWQI) is calculated, followed by preprocessing the dataset for training Artificial Neural Network (ANN) models with cross-validation to ensure accuracy. Results are analyzed annually and seasonally. Feature importance and sensitivity analyses were applied to identify the parameters most influencing developed WQIs in the Gharbia Main Drain. The proposed Artificial Neural Network (ANN) models proved to be highly reliable, effectively capturing the nonlinear relationships within the data and providing more accurate predictions compared to traditional models for WQI evaluation. Five WQI models were developed to evaluate the water quality of the Gharbia Main drain based on different pollution types: biological, industrial, aquatic, agricultural, and overall pollution. Biological pollution emerged as the dominant contributor to poor water quality in the drain. Water quality improved progressively from the inlet to the outfall of the drain, with the Biological Water Quality Index (BWQI) at the Seegaya branch drain (MG02), located at the inlet, increasing only slightly from 8% in 2000–2005 to 23% in 2017–2023. In contrast, the New Gharbia Outfall branch drain (MG14) exhibited better performance, with BWQI improving from 43% to 50% over the same period. Furthermore, water from branch drains located in the downstream reach of the Gharbia Drain consistently showed better water quality than upstream branch drains, indicating that these downstream branches may be suitable for reuse by mixing it with nearby canals before discharging into the main drain.Seasonal analysis highlighted spring and summer as the best seasons at most sites. For example, at MG12, spring BWQI rose from 51% to 92%, the highest observed value. The results of the feature importance analysis identified seven key parameters: Fecal coliform (FC), Biological Oxygen Demand (BOD), Chemical Oxygen Demand (COD), Dissolved Oxygen (DO), Ammonium (NH_4), Total Dissolved Solids (TDS), and Electrical Conductivity (EC) as the most influential when evaluating WQIs. S Sensitivity analysis further supported these findings, showing that variations in these same parameters had the greatest impact on predicted WQIs.

Keywords: *Gharbia Main Drain, Water Quality Index (WQI), Artificial Neural Networks (ANNs), Sensitivity Analysis, Water Management*

1. Introduction

Water is the essence of life and is essential for all activities on Earth, including municipal, agricultural, and industrial processes. However, water scarcity has emerged as a critical global issue, particularly in arid and semi-arid regions. Egypt faces significant challenges in managing its limited water resources while maintaining acceptable water quality. According to the 1959 Nile Water Agreement with Sudan, Egypt is allocated 55.5 billion cubic meters (BCM) of water annually, which has historically supported domestic water supply, irrigation, industrial activities, fisheries, and recreation.

Continuous population increase, land reclamation, and industrial activities have significantly affected these water resources, raising serious concerns about both water quantity and quality (El-Sayed A 2019), (Abd-Elfattah et al., 2021). As a result, Egypt's per capita freshwater availability declined from approximately 1,893 m^3/year in 1959 to 700 m^3/year in 2012, with further reductions projected to reach 505 m^3/year by 2025, placing Egypt well below the internationally recognized water scarcity threshold of 1,000 m^3/year (Esraa et al. 2023).

Accordingly, it has become vital to explore non-conventional water resources. Agricultural drainage water reuse, in particular, has emerged as a crucial source of irrigation water in Egypt. Among these resources, the Gharbia Main Drain plays a significant role, contributing 1.9 billion cubic meters of water annually (El-Sherbiny EK, El-Kassas H 2018). It collects agricultural runoff, domestic wastewater, and industrial effluents, which have negatively impacted its water quality. Studying the water quality in this drain is essential for industry, the environment, and public health protection. It is also critical for the agricultural sector, as it affects agricultural productivity and prevents crop contamination.(Osman et al, 2023; Radwan et al., 2019).

The degradation of water quality in the Gharbia Main Drain is primarily driven by diverse pollution sources. Agricultural runoff introduces significant amounts of nutrients such as nitrogen and phosphorus into the water, leading to eutrophication and harmful algal blooms (Abdelrazek S. 2019; K. 2016). Industrial discharges contribute to hazardous substances, including heavy metals like Lead, Iron, Cadmium, and Nickel, posing serious risks to aquatic life and human health (S. 2015; Vardhan KH, Kumar PS 2019). Additionally, the discharge of untreated or inadequately treated domestic wastewater can lead to significant environmental contamination and pose serious health risks to surrounding communities (Amin MA 2002; MM. 2014).

These pollutants contaminate the water of the Gharbia Drain and threaten its suitability for agricultural irrigation and industrial purposes, posing long-term risks to the region's water resources (Hamed MA., 2019; Mbeche GO.,2021).To effectively monitor, assess, and manage water quality in such complex ecosystems, Water Quality Indices (WQIs) have long been used as critical tools. They summarize the overall quality of water into a single numerical value that reflects the combined impact of multiple water quality parameters, making it easier to understand and communicate the state of water quality (El-Sayed A 2019), (El-Amier et al., 2021), (W.O et al. 2025).

However, traditional methods of calculating WQIs often rely on linear models that may not fully capture the complex, nonlinear relationships among various water quality parameters. These methods also struggle to account for seasonal and environmental changes. These methods often lose important details about individual water quality variables, treat all factors equally, and fail to capture specific local conditions or water uses. Therefore, they may not provide a complete or accurate picture of water quality (Dehkordi DK, 2015; Zhang et al., 2010).

In response to these limitations, there has been a growing interest in applying advanced tools, such as Artificial Neural Networks (ANNs) to develop more accurate and reliable WQIs. ANNs are particularly well-suited for modeling the multifaceted nature of water quality data due to their ability to learn from vast datasets, adapt to new conditions, and predict outcomes with high precision (Tiyasha, Tran MT 2020), (Singh et al., 2009). These models can effectively capture nonlinear relationships between multiple water quality parameters, making them more robust and versatile than traditional linear models (Thakur D 2023), (Gavali KR 2023). ANNs can be trained to predict water quality based on historical data, and they can accommodate new data inputs to update predictions, offering a powerful and prevailing tool for water quality evaluation (Juahir c, 2004; Khadr M, 2017).

Consequently, this study aims to develop more accurate and reliable Water Quality Indices (WQIs) for the Gharbia Main Drain using Artificial Neural Networks (ANNs). The primary objective is to create distinct WQIs for each type of pollution, including biological, industrial, agricultural, and an overall index. The research will involve analyzing extensive water quality data collected over several years and across different seasons, focusing on 19 key parameters representing biological, chemical, and physical indicators. These parameters include: Fecal Coliform (FC), Biological Oxygen Demand (BOD), Chemical Oxygen Demand (COD), Cadmium (Cd), Copper (Cu), Iron (Fe), Manganese (Mn), Zinc (Zn), Nickel (Ni), Lead (Pb), Dissolved Oxygen (DO), Total Suspended Solids (TSS), pH, Electrical Conductivity (EC), Total Dissolved Solids (TDS), Nitrate (NO_3), Ammonium (NH_4), Total Phosphorus (TP), and Total Nitrogen (TN).

The WQIs developed using Artificial Neural Networks (ANNs) will be evaluated by comparing them to the results of the Canadian Water Quality Index (CWQI). Key performance metrics, such as standard deviation (STD), mean square error (MSE), and the coefficient of determination (R^2), will be used to ensure their accuracy, efficiency, and precision. Additionally, sensitivity analysis will evaluate how changes in water quality parameters affect the WQI, while feature importance analysis will highlight the parameters that contribute most significantly to WQI predictions. These approaches may provide valuable insights for environmental monitoring and decision-making in the field of water quality.

This study introduces a novel modeling framework that leverages Artificial Neural Networks (ANNs), integrated with sensitivity analysis and feature importance techniques, to develop data-driven Water Quality Indices (WQIs) for the Gharbia Main Drain. Unlike conventional indices that rely on fixed parameter weightings or focus on single-source pollution, this framework applies adaptive learning across biological, industrial, and agricultural pollution sources simultaneously.

By embedding sensitivity analysis, the model does not only predict water quality but also quantifies how variations in specific pollutants influence overall water quality. In parallel, the feature importance analysis identifies the most critical pollution drivers. This dual predictive and diagnostic capability represents a novel methodological contribution, offering a dynamic, site-specific assessment tool tailored to the challenges of complex drainage systems in arid and semi-arid environments, such as Egypt’s Nile Delta.

The results will offer a powerful tool for assessing and improving water quality in the Gharbia Main Drain, ultimately supporting more sustainable water resource management practices in Egypt’s Nile Delta region. Furthermore, the study’s findings may have broader implications for the application of ANNs in water quality management in other regions facing similar challenges, contributing to global efforts in sustainable water resource management.

2. Material and methods

This section includes a description of the study area, followed by a detailed explanation of the methodology employed in this study. The methodology of this study includes the collection of water quality (WQ) parameters, followed by data entry for the calculation of the Canadian Water Quality Index (CWQI). The dataset is then normalized to facilitate the training of Artificial Neural Network (ANN) models, with cross-validation employed to ensure model accuracy. After achieving the training goals, results are analyzed using two approaches: first, by averaging values over each study period, and second, by calculating the average values of each season (Summer, Fall, Winter, and Spring) to assess seasonal variations.

Finally, sensitivity analysis and feature importance are conducted to identify the key parameters influencing water quality in the Gharbia Main Drain. The framework of the proposed methodology is outlined in **Figure 1**.

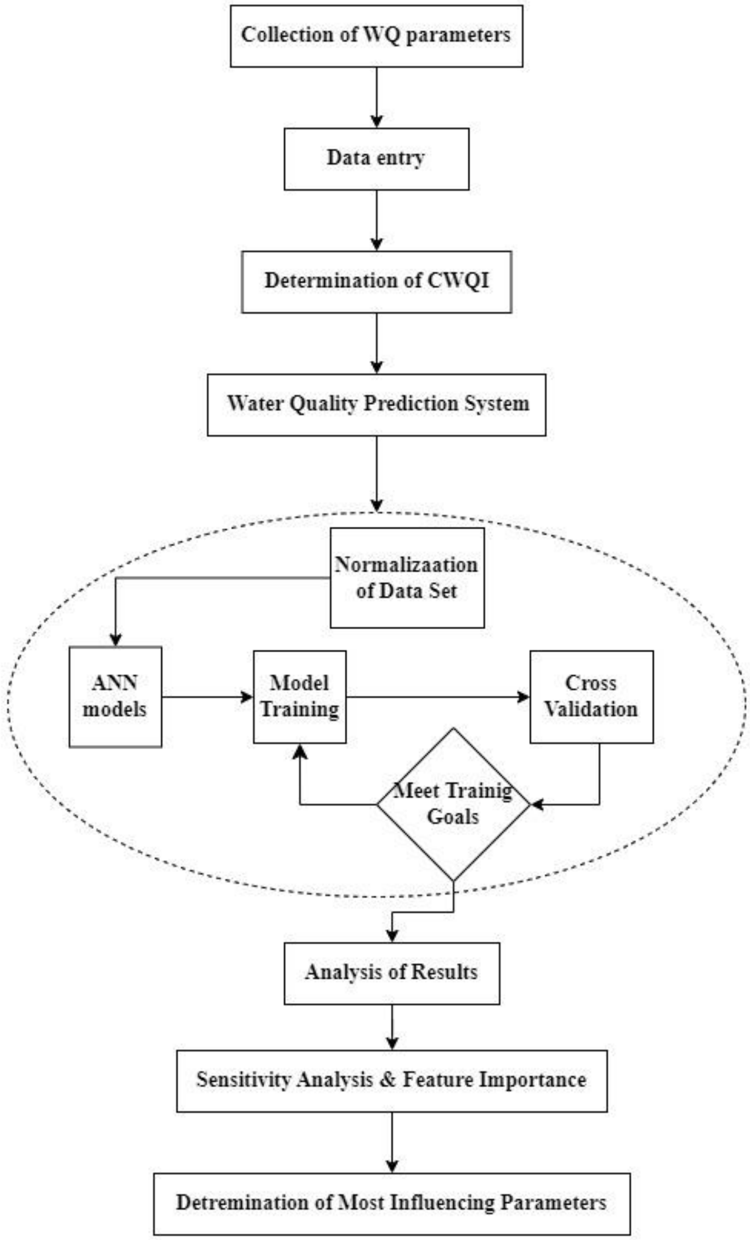


Figure 1: The workflow diagram of the proposed methodology

2.1. Study Area

The Gharbia Main Drain is one of the largest and most critical drainage systems in Egypt, located to the north of Cairo. It extends about 71 kilometers starting at Gharbia governorate and flowing through Dakahilya and Kafr el Sheikh governorates as shown in **Figure 2** (El-Sherbiny EK, El-Kassas H 2018), (Allam et al., 2016). It plays a pivotal role in the region's water management, collecting agricultural drainage water from various sources.

However, the Gharbia Main Drain is also heavily contaminated by multiple pollution sources, which presents significant challenges to water quality and environmental management in the region (Allam et al., 2016).

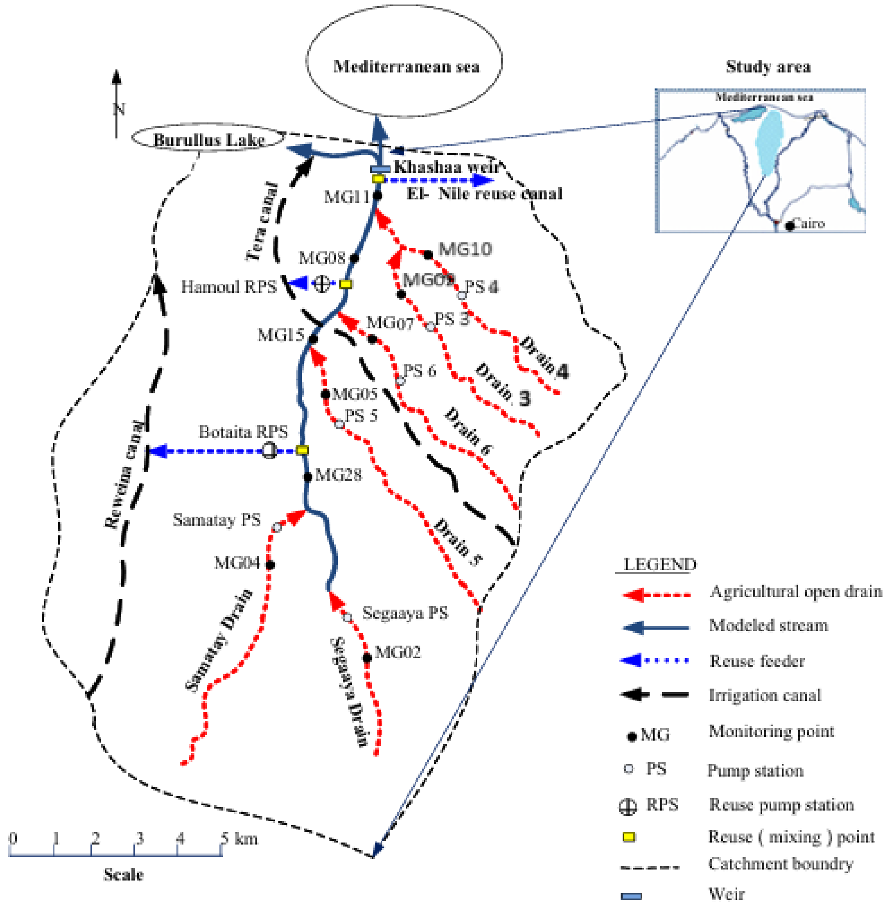


Figure 2: Gharbia's Main Drain Catchment Area after (Allam et al., 2016)

Annually, the Gharbia Main Drain discharges around 1,900 million cubic meters of water, with 50% of this discharge flowing into the Mediterranean Sea. The remaining 50% is reused officially and non-officially for irrigation purposes, affecting public health and agricultural productivity. The drain has 7 main branches, each with a lifting pump station at its outlet. The Gharbia Drain starts from the Segaeaya drain pump station. The other branches are distributed along the Eastern and Western banks of the drain, including Hafir Shehab El-Din, drain No.3, drain No.4, drain No.5, Drain No.6, and Samatay Drain as presented in **Figure 3** (Khalifa AK, Abdel H 2006), (El-Gammal et al., 2009).

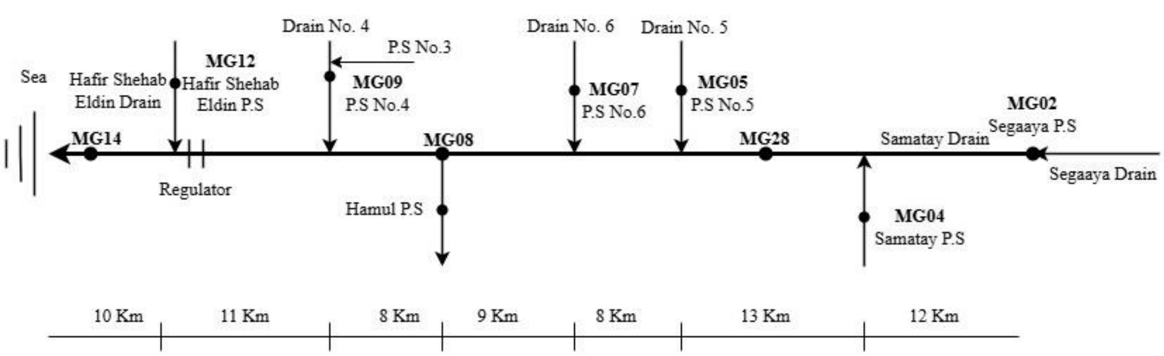


Figure 3: Schematic diagram of the Gharbia Main Drain after (El-Gammal et al., 2009)

2.2. Data Collection

The evaluation of water quality in the Gharbia Main Drain was based on water sample data collected from previous research (Osman et al., 2023),(Abd-Elfattah et al., 2021),(Abosena et al., 2021),(El-Sherbiny EK, El-Kassas H 2018),(El-amier et al. 2017),(Mohamed, Elansary, and Moussa 2017),(Taha et al., 2012),(Zaghloul SS 2011),(El-Gammal et al., 2009),(Donia et al., 2009),(Khalifa AK, Abdel H 2006), and the annual reports of the Drainage Research Institute (DRI) (Drainage Research Institute (DRI) yearbooks 2000:2015), covering the period from **2000 to 2023** for the Gharbia Main Drain and its branches.

During the period from 2000 to 2015, water quality samples were collected monthly at each monitoring site, providing 12 data points per year per site. For the period from 2017 to 2023, sampling frequency shifted to seasonal collection, resulting in 4 data points per year per site. It is important to note that data were missing for the period between July 2015 and January 2017. All collected data underwent quality checks and preprocessing before being used in the ANN models. This preprocessing included reviewing for completeness and consistency, and in cases where minor gaps were identified, these were filled using linear interpolation. Finally, the complete dataset was normalized using min-max scaling to ensure uniform contribution of all parameters during model training.**Table 1** summarizes the average values for each water quality parameter across the study period, providing a baseline characterization of water quality conditions in the Gharbia Main Drain.

Table 1. Average values of key water quality parameters in the Gharbia Main Drain during the study period

parameter	Unit	Stations								
		MG02	**MG04**	**MG28**	**MG05**	**MG07**	**MG08**	**MG09**	**MG12**	**MG14**
FC	MPN/100ml	3219564	968010	1243241	581125	1603589	546260	510443	125692	330423
BOD	mg/L	43.3	34.2	33.5	29.1	30.9	31.9	31.0	22.6	26.4
COD	mg/L	56.6	46.1	44.3	38.0	41.4	42.6	41.6	30.8	33.5
Cd	mg/L	0.001	0.001	0.002	0.001	0.001	0.001	0.001	0.002	0.001
Cu	mg/L	0.065	0.055	0.050	0.057	0.050	0.053	0.080	0.041	0.056
Fe	mg/L	0.682	0.756	0.604	0.671	0.697	0.648	0.680	0.668	0.608
Mn	mg/L	0.2	0.215	0.223	0.2	0.270	0.193	0.219	0.226	0.209
Zn	mg/L	0.035	0.028	0.035	0.029	0.030	0.046	0.034	0.026	0.032
Ni	mg/L	0.003	0.005	0.009	0.005	0.005	0.005	0.006	0.005	0.003
Pb	mg/L	0.007	0.012	0.009	0.012	0.009	0.013	0.009	0.016	0.011
DO	mg/L	1.17	1.56	1.34	5.57	2.06	1.81	2.29	2.90	2.43
TSS	mg/L	65.6	60.2	42.0	46.1	82.4	42.6	46.1	102.6	75.5
pH		7.5	7.5	7.5	7.5	7.5	7.5	7.5	7.6	7.6
EC	dS/m	1.5	1.2	1.2	1.3	3.2	1.5	1.5	5.0	3.8
TDS	mg/L	996.1	854.7	851.6	866.7	2074.7	1014.2	1082.7	3222.0	2463.2
NO_3	mg/L	8.2	7.1	6.9	6.8	11.2	7.4	7.6	11.9	10.6
NH_4	mg/L	4.1	3.3	5.3	3.7	3.4	2.4	2.2	3.7	3.1
TP	mg/L	1.04	0.93	0.92	0.73	0.54	0.75	0.72	0.60	0.70
TN	mg/L	12.7	10.6	12.5	10.9	15.0	10.3	10.2	16.3	14.1

The Central Laboratories for Environmental Quality Monitoring (CLEQM) of the National Water Research Center (NWRC) analyzed these samples, which were collected from various locations along the Gharbia Main Drain and its branches according to the National Water Quality Monitoring Network (NWQM 2003). The coordinates and names of these monitoring sites are presented in **Table 2**.

Table 2. Water Quality Monitoring sites in the Gharbia Main Drain and its branches

Site	Site Name	Co-ordinates	
		Latitude	Longitude
MGO2	Segaaya Pump Station(P.S.)	31.018816	31.071283
MG04	Samatay P.S.	31.121244	31.055928
MG05	P.S No. 5	31.183316	31.125833
MG07	P.SNo. 6	31.287333	31.13875
MG08	El Hamul P.S	31.309166	31.141166
MG09	P.SNo. 4	31.303250	31.197733
MG12	Hafir Shehab El Din P.S	31.490766	31.151633
MG14	New Gharbia Outfall	31.49755	31.1469
MG28	Bridge Down Stream Samaty P.S.	31.14258	31.0542

2.3. Water Quality Parameters and Data Categorization

The previous nineteen water quality parameters mentioned in the introduction were categorized into four distinct groups based on their pollution sources: biological, industrial, agricultural, and overall pollution. In addition to these specific categories, an aquatic category was created to evaluate aquatic life along the main drain and its branches. The comprehensive categories illustrated in **Figure 4** were used to develop a holistic ANN model that integrates data from all pollution sources, providing a complete assessment of water quality.

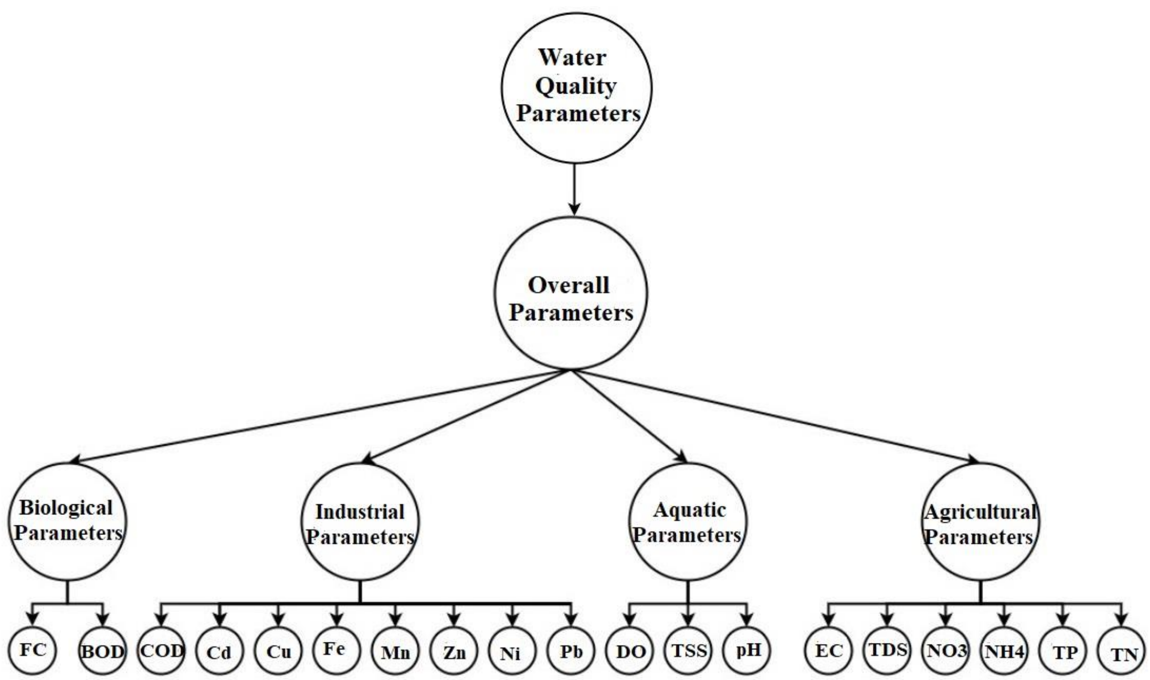


Figure 4: Categorization of Water Quality Parameters

To further contextualize the importance of these parameters, their permissible limits according to Egyptian Decree Law No. 48 of 1982 (*EGYPT Decree Law 48* / 1982) and its Executive Regulations Law No. 92 of 2013(Egypt Decree 2013), , along with the Food and Agriculture Organization (FAO) guidelines for irrigation water quality (FAO, 2003), are summarized in **Table 3**. These national standards set the maximum allowable concentrations for key water quality parameters in drainage water prior to discharge into freshwater bodies, serving as an important regulatory benchmark for evaluating water quality in the Gharbia Main Drain, complementing the scientific evaluation conducted using the Canadian Water Quality Index (CWQI).

Table 3. Selected Water Quality Parameters and their Guidelines in Drainage Water

Water Quality Parameter	Unit	Guideline Decree-Law No. 92 of 2013	FAO Guideline
FC	MPN/100ml	5000	1000
BOD	mg/L	30	
COD	mg/L	50	
Cd	mg/L	0.03	0.01
Cu	mg/L	1	0.2
Fe	mg/L	3	5
Mn	mg/L	2	0.2
Zn	mg/L	2	2
Ni	mg/L	0.1	
Pb	mg/L	0.1	5
DO	mg/L	>5	
TSS	mg/L	500	
pH		6.5:8.5	6.5:8.4
EC	dS/m	0.64	3
TDS	mg/L	1000	2000
NO_3	mg/L	45	5:30
NH_4	mg/L	0.5	5
TP	mg/L	3	
TN	mg/L	15	

2.4. Canadian Water Quality Index (CWQI)

The Canadian Water Quality Index (CWQI), developed by the Canadian Council of Ministers of the Environment (CCME), simplifies water quality data by combining multiple measurements into a single score. This standardized method provides a clear summary of water quality for experts and the public, enabling easy communication and trend monitoring over time(Tyagi et al., 2013),(Canadian Council of Ministers of the Environment. 2001).

The CWQI serves as the target output for the ANN models. By training the ANN on the input water quality parameters, the model learns to predict the CWQI. The parameters selected for CWQI calculation were chosen to align with the permissible limits set by Decree Law No. 48 of 1982 and its Executive Regulations, Decree No. 92 of 2013, as shown in **Table 3**. WQIs for the Gharbia Main drain were calculated according to the mentioned data categorization in the previous section **Figure 4**.

2.4.1. Calculation of CWQI

The CWQI is calculated using three key factors $\mathbf{F_1}$, $\mathbf{F_2}$, $\mathbf{F_3}$ as the following:

Scope (F_1): Reflects the extent of non-compliance with water quality guidelines over a given period.

$$F_1 = \left(\frac{\text{Number of failed variables}}{\text{Total number of variables}} \right) \times 100 \quad (1)$$

Frequency (F_2): Indicates the proportion of individual tests that fail to meet water quality objectives.

$$F_2 = (\text{tests}) \times 100 \quad (2)$$

Amplitude (F_3): Quantifies how the failed test values deviate from their respective objectives. The calculation of F_3 involves three steps:

1. Excursions Calculation:

For test values that should not exceed the objective:

$$\text{Excursion} = \frac{\text{Failed Test Value} - \text{Objective}}{\text{Objective}} \tag{3}$$

For test values that should not fall below the objective:

$$\text{Excursion} = \frac{\text{Failed Test Value} - \text{Objective}}{\text{Failed Test Value}} \tag{4}$$

2. Normalized Sum of Excursions (NSE):

Obtained by summing all the excursions and dividing by the total number of tests:

$$\text{NSE} = \frac{\sum \text{excursions}}{\text{Total number of tests}} \tag{5}$$

3. F_3 Calculation:

The amplitude factor is calculated using an asymptotic function that scales the normalized sum of excursions (NSE) to a range between 0 and 100:

$$F_3 = \left(\frac{\text{NSE}}{0.01\ \text{nse} + 0.01} \right) \tag{6}$$

Once F_1, F_2, and F_3 have been calculated, the **CWQI** is determined by combining these factors into a final index score:

$$\text{CWQI} = 100 - \left(\frac{\sqrt{F_1^2 + F_2^2 + F_3^2}}{1.732} \right) \tag{7}$$

where 1.732 is a correction factor.

2.4.2. CWQI Score Categorization

The CWQI score is categorized into five levels to describe the water quality, as summarized in **Table 4** (Tyagi et al., 2013).

Table 4. Canadian Water Quality Index (CWQI) Score Categories

CWQI Score	Water Quality Index Category
95-100	Excellent
80-94	Good
65-79	Fair
45-64	Marginal
0-44	Poor

2.5. Artificial Neural Networks (ANNs)

An artificial neural network (ANN) is a highly parallel and distributed information processing system modeled after the neural networks in the human brain (Altunkaynak A. 2007). It is a powerful soft computing technique used for linear and nonlinear approximations across various fields (Kişi Ö. 2006), which traditional statistical methods may not effectively capture (Tyagi et al., 2013). The ANN consists of an array of interconnected neurons designed to solve specialized problems, where all input nodes feed into the first hidden layer, the hidden layers pass information sequentially to each other, and the final hidden layer feeds into the output layer (Konaté et al., 2015). The architecture of an ANN includes a specified number of hidden layers, along with neurons distributed across the input layer, hidden layers, and output layer.

The ANN models developed in this study were structured to optimize performance across different water quality categories. The input layer consisted of nodes representing various water quality parameters, while multiple architectures of hidden layers were explored to find the optimal configuration, as the relation between inputs and outputs is nonlinear, as shown in **Figure 5** for biological parameters.

The tested configurations included one hidden layer with 5, 10, and 15 neurons; two hidden layers with 5, 10, and 15 neurons in each layer; and three hidden layers, each with 10 neurons. The output layer predicted the Water Quality Index (WQI), with the Canadian Water Quality Index (CWQI) serving as the target variable.

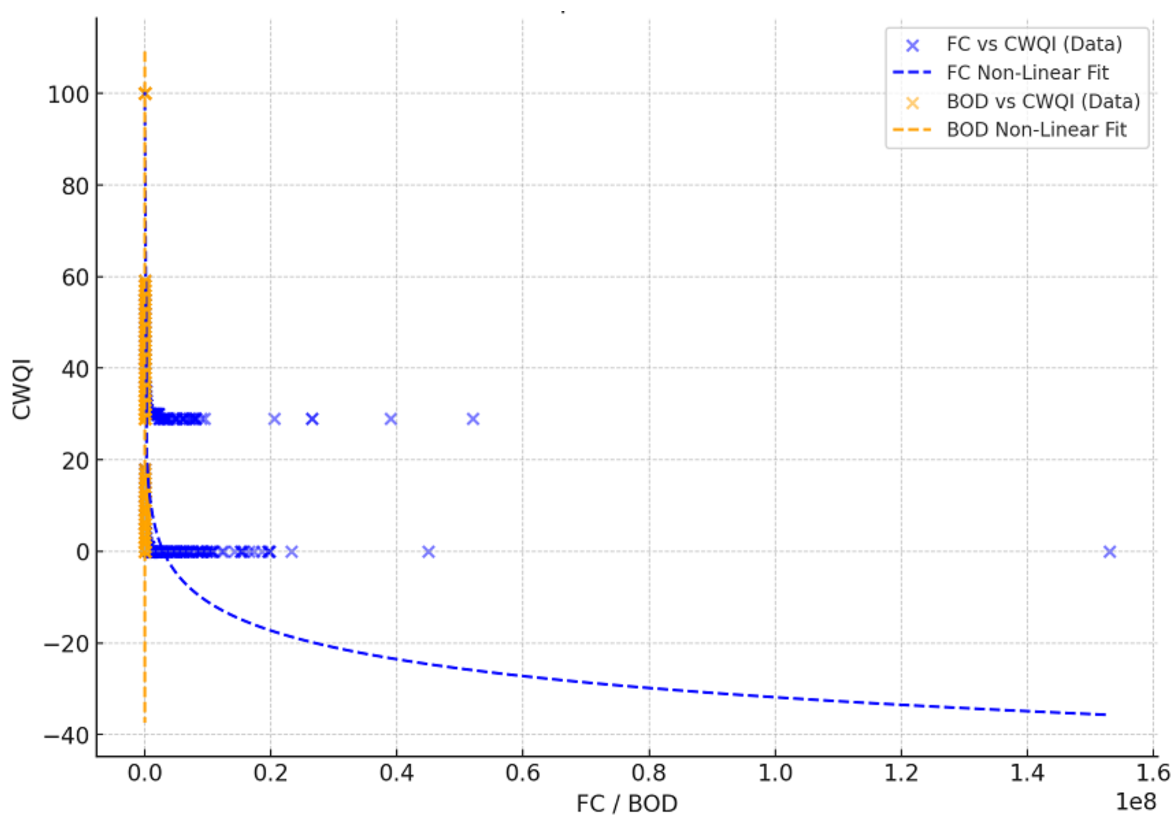


Figure 5: Visualization of the Nonlinearity of Biological Data

To prevent overfitting and ensure robust generalization, the dataset, which included 2,016 water quality records collected from nine monitoring sites over the period from 2000 to 2023, was divided into three subsets: 70% for training, 15% for validation, and 15% for testing. This approach allowed the model to learn effectively while retaining an independent dataset for unbiased evaluation. Additionally, early stopping was applied during training, terminating the process when validation error began to increase, preventing unnecessary overfitting. For model training, the ANN used a backpropagation algorithm, which updates the network's weights to minimize prediction errors. The sigmoid activation function was applied in the hidden layers, transforming the inputs into outputs within a range of 0 to 1 (Abdel-Fattah MK, Mokhtar A 2021):

$$f(x) = \frac{1}{1+e^{-x}} \tag{8}$$

In forward propagation, the output of each neuron y_j was computed by summing the weighted inputs and applying the activation function:

$$y_j = f\left(\sum_{i=1}^{n} w_{ij}\, x_i \; + \; b_j\right) \tag{9}$$

where y_j is the output of neuron j, w_{ij} are the weights, x_i are the inputs, and b_j is the bias term.

During backpropagation, the model's weights were updated to minimize the error function, typically the mean squared error (MSE). The weight update rule followed this equation:

$$w_{ij}(t+1) = w_{ij}(t) - \eta \frac{\partial E}{\partial w_{ij}} \tag{10}$$

where $w_{ij}(t)$is the weight at iteration t, η is the learning rate, and E is the error function.

The performance of the ANN models was evaluated using three key metrics (Abdel-Fattah MK, Mokhtar A 2021): **standard deviation (STD)**, **mean squared error (MSE)**, and the **coefficient of determination (R^2)**. A lower standard deviation (STD) indicates more consistent predictions and is calculated as:

$$STD = \sqrt{\frac{1}{n}\sum(y_k - \tilde{y})^2} \tag{11}$$

where y_k represents the predicted value, and ỹ denotes the mean of the predicted values.

The mean squared error (MSE) reflects the model's accuracy by quantifying the average squared difference between actual and predicted values and is computed using the formula:

$$MSE = \frac{1}{n}\sum_{k=1}^{n}(y_k - \tilde{y})^2 \tag{12}$$

Where y_k is the actual output, and ỹ is the predicted output. A lower MSE indicates greater accuracy and better error minimization in the model's predictions.

The coefficient of determination (R^2) measures the proportion of variance in the dependent variable that can be explained by the independent variables and is calculated as:

$$R^2 = 1 - \frac{\sum_{k=1}^{n}(y_k - \tilde{y}_k)^2}{\sum_{k=1}^{n}(y_k - \tilde{y})^2} \tag{13}$$

Where y_k denotes the actual output, $\tilde{y}_k$ represents the predicted output, and ỹ is the mean of the actual values. An R^2 value close to 1 signifies strong predictive performance, indicating that the model effectively captures the variance in the data.

The final models shown in **Figure (6a–6f)** were selected based on achieving the lowest STD and MSE values, along with the highest R^2, ensuring both accuracy and reliability in predicting water quality.

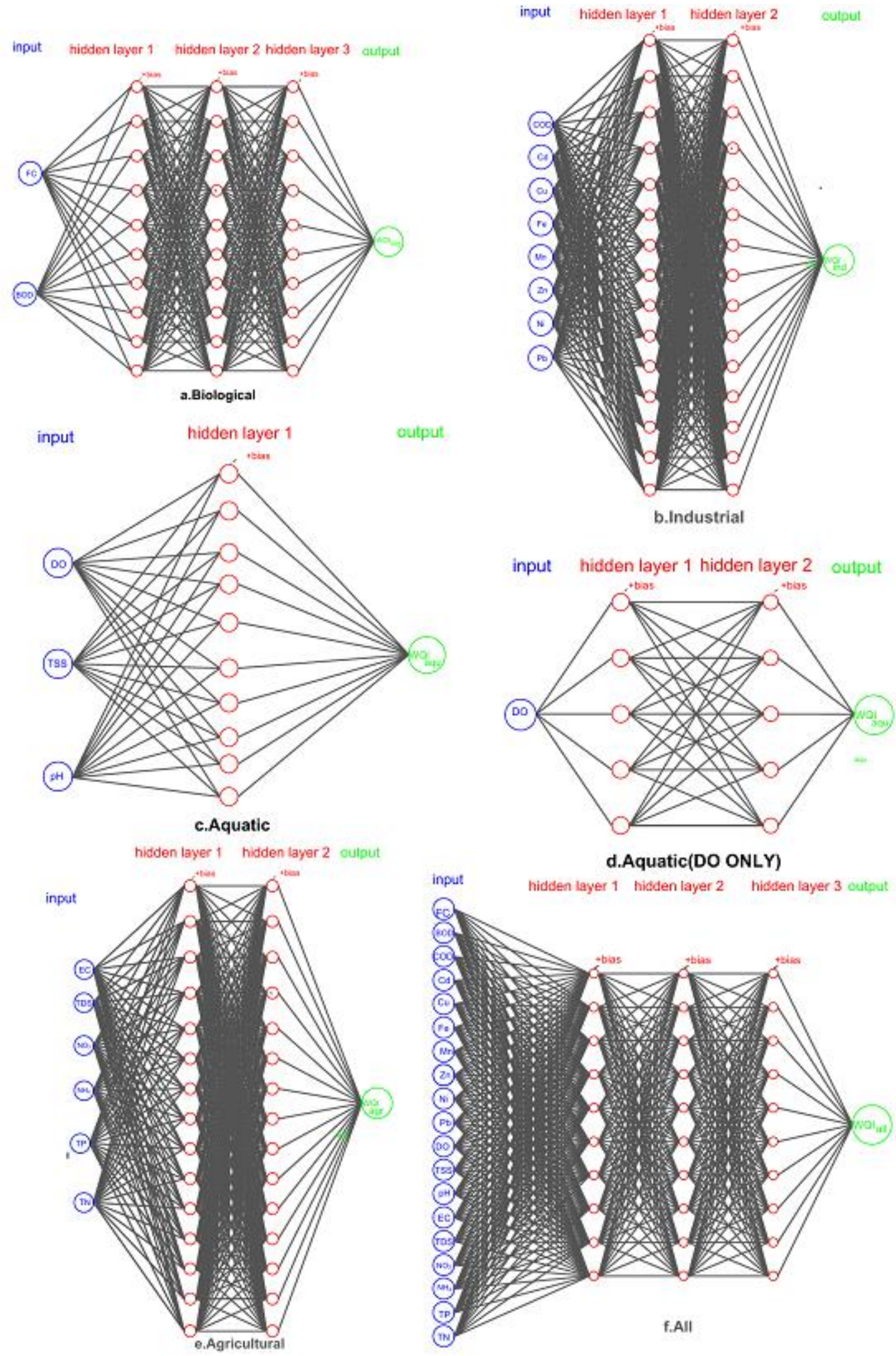


Figure (6a–6f): Architecture of Best Performing ANN models

2.6. Calculation of Feature Importance for the Most Influencing Parameters

A Random Forest regression approach was employed to assess the relative importance of predictor variables in influencing the response variable(L. 2001). This method is particularly effective in handling non-linear relationships and providing insights into the contribution of each feature. A Random Forest model, consisting of a specified number of trees (e.g., 100), was trained using the predictor matrix **X** and the response variable **Y**. Model performance was evaluated using the Out-of-Bag (OOB) error, which provides an unbiased estimate of the prediction error.

A permutation-based feature importance method was used to quantify each feature's importance. This involved permuting the values of each feature and measuring the resulting increase in OOB error. The feature importance FI_j for each feature j was calculated as the average difference between the original OOB error and the OOB error after permuting the feature's values:

$$FI_j = \frac{1}{T}\sum_{t=1}^{T}\left(OOB\ Error_{perm.j}^{t} - OOB\ Error_{orig}^{t}\right) \quad (14)$$

Where; T is the total number of trees, $OOB\ Error_{perm.j}^{t}$ is the OOB error after permuting

feature j and $OOB\ Error_{orig}^{t}$ is the original OOB error.

The feature importance scores were then normalized to provide a probabilistic interpretation. This normalization was performed by dividing each feature's importance score by the sum of the importance scores for all features:

$$NFI_j = \frac{FI_j}{\sum_{k=1}^{p} FI_k} \quad (15)$$

Where p represents the total number of features. This process allowed for the identification of the most significant predictor variables, providing valuable insights into their relative contributions to predicting the response variable.

2.7. Sensitivity Analysis

Sensitivity analysis was performed across various predictive models to evaluate the impact of changes in key input variables on the CWQI (Ibrahim J, Chen MH 2009). The process began with the normalization of input variables to standardize their scales, ensuring all the models could handle them effectively. Several models, including Linear Regression, Polynomial Regression, Decision Trees, Random Forests, and Neural Networks, were trained and validated using k-fold cross-validation (k = 5). The best-performing model was selected based on its Root Mean Squared Error (RMSE) and coefficient of determination (R^2) During the sensitivity analysis, key input variables were varied systematically within a range of -50% to +50% to evaluate their impact on the predicted CWQI. The model's predictions were recorded for each variation, and the mean CWQI was calculated, allowing the identification of the most influential variables in predicting water quality.

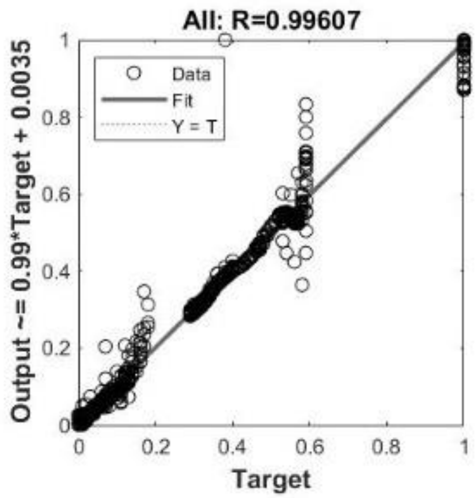
3. Results and Discussion

3.1. Artificial Neural Network Models for Estimating WQIs

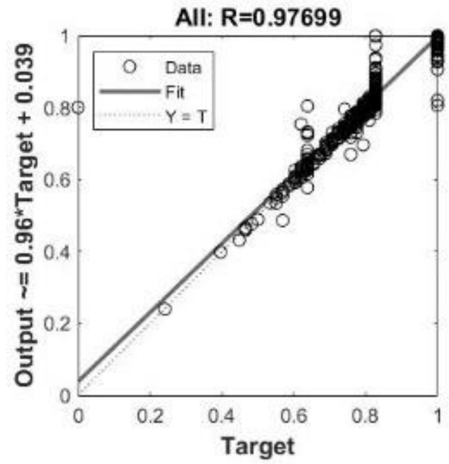
The best-performing neural network structures for each category are shown in **Figure (6a–6f)**, with standard deviation (STD), mean squared error (MSE), and coefficient of determination (R^2) summarized in **Table 5**. The coefficient of determination (R^2) presented in **Figure (7a–7f)** exceeds 0.97 indicating a strong correlation between observed Water Quality Index (WQI) values from traditional methods and those predicted by the ANN models.

Table 5. Summary of Best-Performing Neural Network (NN) Models for Different Models

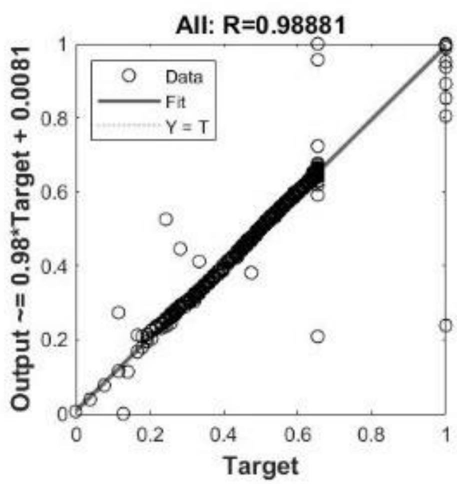
Models	**STD**	**MSE**	**R^2**
Biological model	0.0263	0.007	0.99607
Industrial model	0.0248	0.00061	0.97699
Aquatic model	0.0279	0.00078	0.9881
Aquatic (DO only)	0.0027	0.000007	0.99995
Agricultural model	0.0246	0.0006	0.99232
Overall model	0.0242	0.00058	0.99417



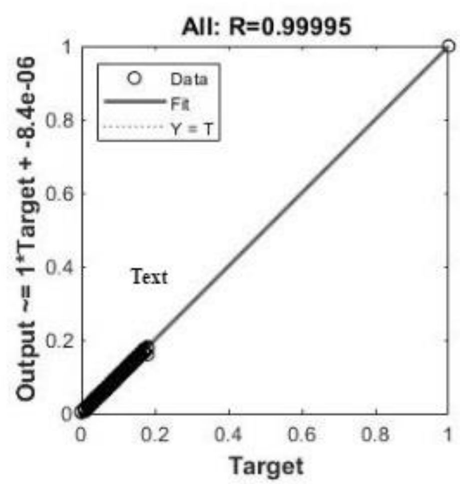
a. Biological Model



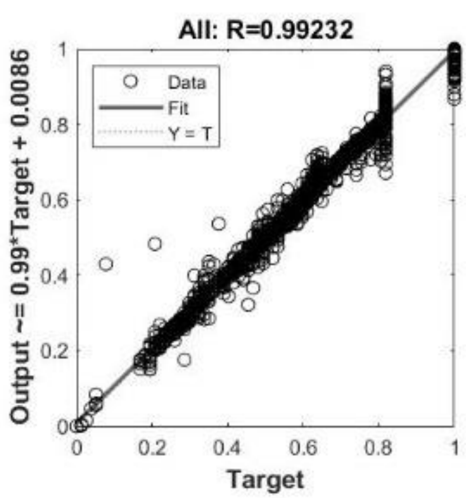
b. Industrial Model



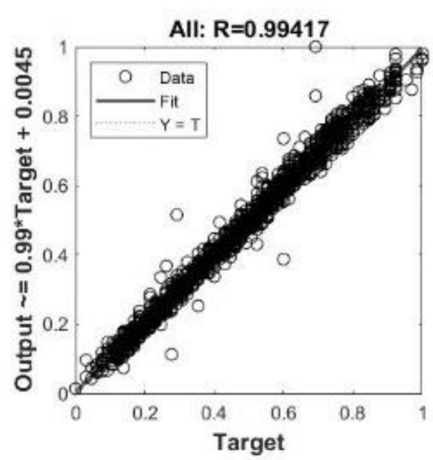
c. Aquatic Model



d. Aquatic Model (DO only)



e. Agricultural Model



f. Over All Model

Fig. (7a–7f) Comparison of the Actual WQI and the Model Predictions

3.2. Artificial Neural Networks Parameters Estimation

Table (6a-6d) summarizes the weight parameters of the artificial neural network for the biological model, detailing connections from inputs to successive hidden layers (H1, H2, H3) and ultimately to the output layer.

Table 6. a. Initial Input-to-Hidden Layer 1 Weight Parameters for Biological Model

Predictors	**Predicted**									
	Hidden Layer (1)									
	H(1,1)	**H(1,2)**	**H(1,3)**	**H(1,4)**	**H(1,5)**	**H(1,6)**	**H(1,7)**	**H(1,8)**	**H(1,9)**	**H(1,10)**
FC	-4.04	1.74	3.92	3.88	0.09	0.72	-0.41	21.09	3.13	3.87
BOD	-1.80	4.07	2.06	-2.33	4.53	-4.35	4.38	-0.01	-10.3	-2.61

Table 6.b. Hidden Layer 1 to Hidden Layer 2 Weight Parameters for Biological Model

Predictors	**Predicted**									
	Hidden Layer (2)									
	H(2,1)	**H(2,2)**	**H(2,3)**	**H(2,4)**	**H(2,5)**	**H(2,6)**	**H(2,7)**	**H(2,8)**	**H(2,9)**	**H(2,10)**
H(1,1)	-0.56	0.47	-0.29	-1.05	0.76	0.45	-0.67	-0.28	-0.73	-0.36
H(1,2)	0.31	0.94	0.38	0.35	0.42	0.95	-0.92	-0.15	-0.94	-0.59
H(1,3)	0.21	0.09	0.57	-0.78	-0.66	-0.43	-0.80	0.40	-1.11	1.03
H(1,4)	-0.55	0.004	-0.73	0.27	1.18	-0.73	-0.30	0.51	-0.34	-0.48
H(1,5)	-0.34	-0.62	-0.48	-0.05	-0.68	0.35	-0.24	-1.05	0.24	0.49
H(1,6)	0.07	0.33	-0.33	-0.36	-0.10	0.37	0.54	-0.62	-0.05	0.18
H(1,7)	-0.84	0.40	0.02	0.21	0.93	0.07	-0.61	0.43	0.21	-0.08
H(1,8)	0.85	-0.52	-14.2	-20.6	-1.81	16.87	0.48	1.44	-0.64	0.78
H(1,9)	-0.57	-0.25	-1.50	-0.05	6.09	-1.40	-5.42	-0.76	4.51	0.05
H(1,10)	0.59	-0.77	0.04	0.65	-0.54	0.37	-0.62	-0.39	-0.06	-0.73

Table 6.c. Hidden Layer 2 to Hidden Layer 3 Weight Parameters

Predictors	**Predicted**									
	Hidden Layer (3)									
	H(3,1)	**H(3,2)**	**H(3,3)**	**H(3,4)**	**H(3,5)**	**H(3,6)**	**H(3,7)**	**H(3,8)**	**H(3,9)**	**H(3,10)**
H(2,1)	-0.20	-1.07	0.12	-0.84	0.10	-0.06	-0.20	0.62	0.04	0.06
H(2,2)	-0.30	-0.05	-0.35	0.33	-0.27	-0.91	0.66	0.78	0.31	-0.29
H(2,3)	0.11	-0.15	0.59	3.90	-0.92	0.88	-0.58	4.94	-0.12	-12.6
H(2,4)	0.62	-0.44	0.57	6.84	-0.51	-0.70	-0.33	6.82	-0.94	-18.1
H(2,5)	0.22	1.29	-0.14	1.57	0.18	-4.71	-1.89	0.24	1.33	-2.46
H(2,6)	0.29	0.23	1.05	-4.78	1.85	1.66	-0.54	-5.88	0.04	15.37
H(2,7)	0.64	1.28	0.62	-0.01	-0.80	3.90	0.74	-0.59	1.05	-0.96
H(2,8)	-0.75	-0.97	-1.03	0.34	-1.18	0.72	0.70	0.33	-0.28	-1.04
H(2,9)	-0.38	-0.78	-0.37	-1.42	-0.41	-3.86	-0.35	-0.61	0.53	1.76
H(2,10)	-1.03	-0.58	0.42	-0.72	-0.61	-0.64	-0.48	-0.04	-0.66	-0.88

Table 6.d. Output Weights from Hidden Layer 3 for Biological Model

Predictors	Output
H(3,1)	0.39
H(3,2)	2.07
H(3,3)	-0.23
H(3,4)	-6.80
H(3,5)	1.77
H(3,6)	-0.79
H(3,7)	-1.35
H(3,8)	2.60
H(3,9)	-1.31
H(3,10)	-10.79

The weight matrices presented in **Tables (6a - 6d)** represent the internal structure and learning process of the Artificial Neural Network (ANN) developed for predicting the Biological Water Quality Index (BWQI). These weights define how inputs (water quality parameters) influence successive hidden layers and ultimately contribute to the predicted BWQI at the output layer. The strong initial weights between the input parameters, particularly Fecal Coliform (FC) and Biological Oxygen Demand (BOD), and the first hidden layer neurons indicate that these two parameters play a dominant role in shaping the predicted BWQI. As the signal propagates through deeper layers, these influences are refined, capturing complex interactions between biological pollutants.

3.3. WQIs results for Different Pollution Types along the period 2000-2023 in Gharbia Main Drai

The analysis of WQI results for the different studied categories shows that the trend of WQIs is the same across all water quality stations in the Gharbia Main Drain and its branches. The results for MG08 are shown in the given **Figure 8.** The Industrial model ranges between 92 and 100, placing it in the Good to Excellent category of water quality. The Agricultural model shows values between 59 and 75, which fall within the Marginal to Fair category. The overall Water Quality Index (WQI) for this site ranges between 54 and 69, indicating a Marginal to Fair water quality status. In contrast, the Biological model exhibits values ranging from 18 to 47, placing it in the Poor category. Similarly, the Aquatic model ranges from 10 to 19, which also falls into the Poor category.

As a result, Biological pollution is the dominant concern, as the lowest values of WQI were observed in both the Biological and Aquatic categories. This highlights that the most significant impact on water quality in the Gharbia Drain comes from biological pollutants.

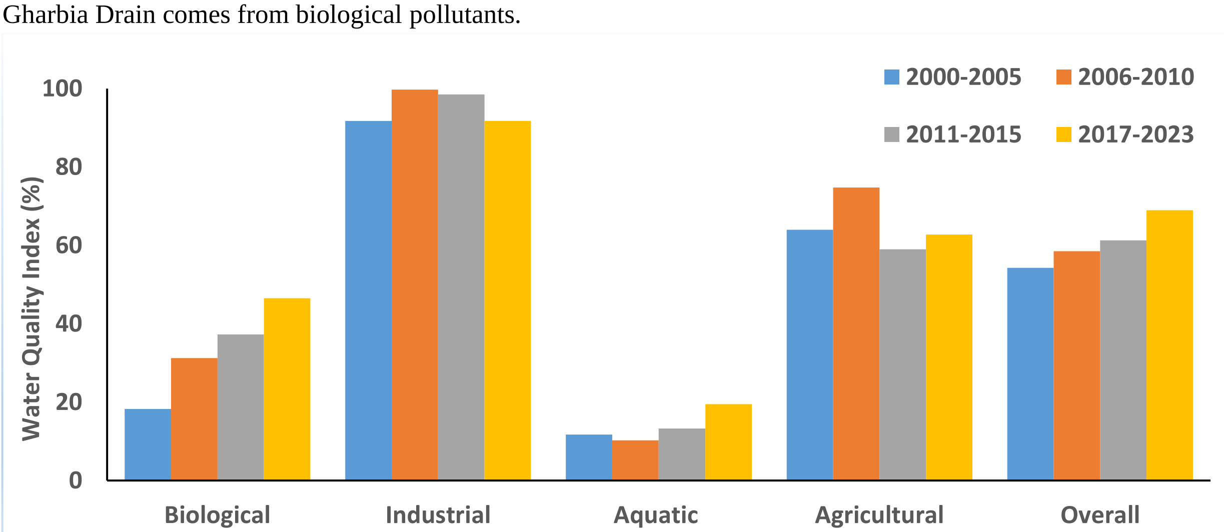


Figure 8: WQIs for Different Categories at MG08 (2000-2023)

3.4. Sites Analysis for Annual Biological Pollution in the Gharbia Main Drain (2000-2023)

The analysis of the Biological Water Quality Index (BWQI) from 2000 to 2023, as shown in **Figure 9**, reveals significant variations in pollution levels across key monitoring sites along the Gharbia Main Drain: MG02, MG28, MG08, and MG14. Among these sites, MG14 exhibited the highest biological water quality throughout the study period, with its BWQI increasing from 43 in 2000–2005 to 50 in 2017–2023. MG08 showed a clear upward trend, rising from 18 in 2000–2005 to 47 in 2017–2023. Similarly, MG28 experienced a gradual increase, with BWQI improving from 16 to 34 over the same period. In contrast, MG02, located at the inlet of the drain, consistently recorded the lowest biological water quality. Its BWQI increased only slightly, from 8 in 2000–2005 to 23 in 2017–2023. Overall, the data demonstrate that MG14 stands out as the least polluted site, while MG02 remains the most polluted throughout the monitoring period.

This analysis demonstrates that biological pollution is highest at the inlet of the drain and gradually improves toward the outfall. Additionally, there has been a noticeable reduction in biological pollution with each successive period, which aligns with the establishment of wastewater treatment plants in the catchment area of the Gharbia Main Drain by the government in recent years.

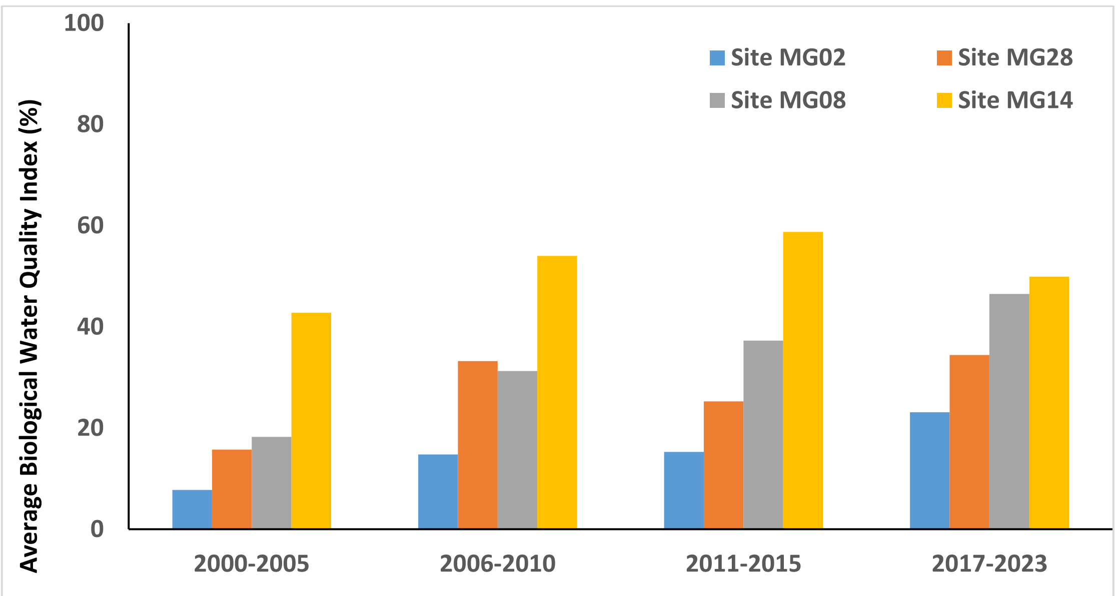


Figure 9: Average BWQI across Main Sites (2000-2023)

3.5. Seasonal Variations in Biological Pollution in the Gharbia Main Drain (2000-2023)

The seasonal variations in water quality across the monitoring sites of the Gharbia Main Drain reveal distinct trends over the periods analyzed as shown in **Fig. (10a–10d)**. At MG02, spring and summer consistently exhibited the highest biological water quality across the monitored seasons. During summer, BWQI increased from 6 in 2000–2005 to 33 in 2017–2023. Spring BWQI followed a fluctuating pattern, starting at 11 in 2000–2005, rising to 19 in 2006–2010, then dropping to 10 in 2011–2015, before improving again to 24 in 2017–2023. Winter and fall consistently showed lower biological water quality, with fall BWQI increasing from 5 to 23, and winter BWQI starting at 9, peaking at 13 in 2011–2015, and then slightly declining to 12 in 2017–2023.

At MG28, winter exhibited the greatest improvement, with BWQI rising from 10 in 2000–2005 to 45 in 2017–2023. Spring also showed a steady increase, improving from 14 to 38 over the same period. Summer BWQI remained relatively stable, starting at 26 and fluctuating slightly to 31 in the final period. Fall initially improved from 13 to 34 by 2006–2010, but then dropped to 16 in 2011–2015, before recovering to 24 in 2017–2023.

MG08 shows the most significant improvements in summer and spring. Spring consistently showed the highest biological water quality, with BWQI rising from 14 in 2000–2005 to 57 in 2017–2023. Summer followed a similar improvement trend, increasing from 34 to 51 over the same period. Fall and winter remain the worst, , with fall BWQI rising from 16 to 41, and winter BWQI increasing from 9 to 37.

At MG14, both summer and spring are the best seasons. Spring initially had the highest BWQI, rising from 38 in 2000–2005 to a peak of 86 in 2011–2015, but then declining to 44 in 2017–2023. Summer BWQI dropped from 77 in 2000–2005 to 52 in 2006–2010 and remained around the same level in the later periods, ending at 51. While winter and fall remain the worst.

These seasonal trends indicate that water quality improvements have occurred across all sites, although some fluctuations remain, particularly in spring and summer at some locations. The overall seasonal pattern highlights that biological water quality tends to improve more in spring and summer, while fall and winter often reflect poorer conditions.

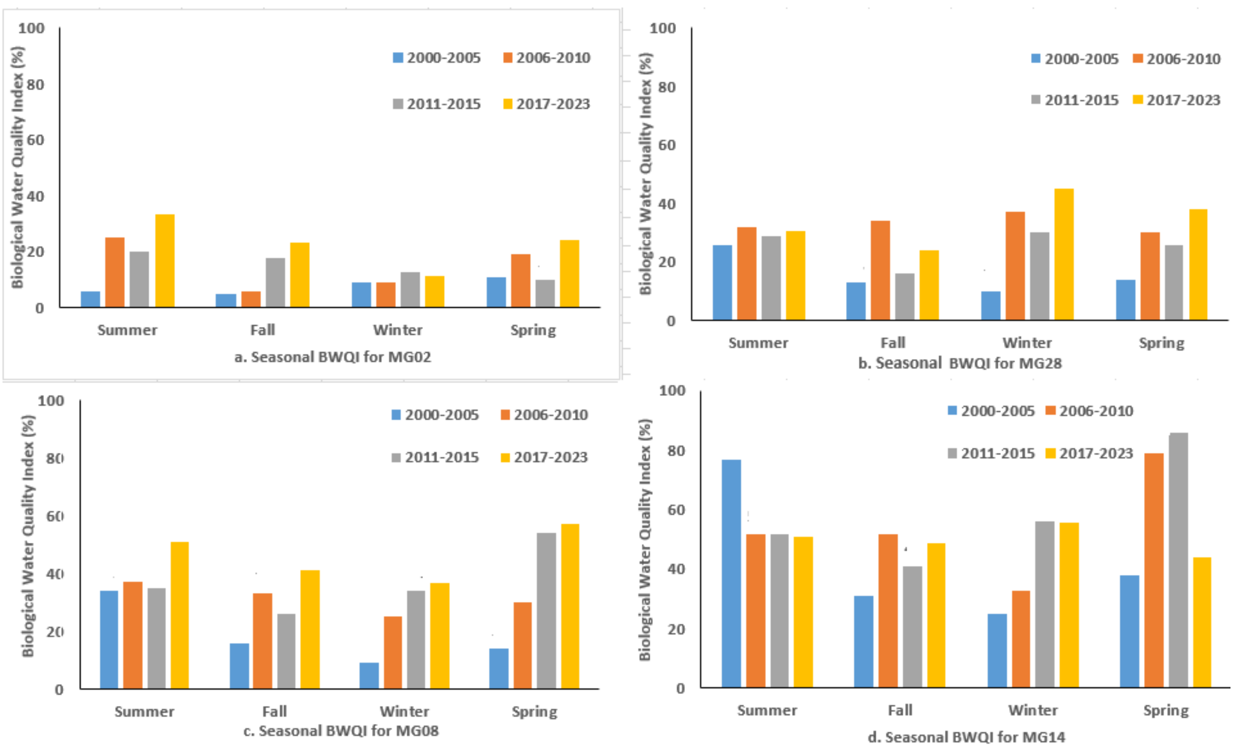


Figure (10a-10d): Seasonal Variation of BWQI Across Main Sites (2000-2023)

3.6. Annual Analysis of Biological Pollution at Branch Sites in the Gharbia Drain (2000-2023)

The analysis of the Biological Water Quality Index (BWQI) from 2000 to 2023, as illustrated in the given **Figure 11**, reveals significant variations in water quality across key monitoring sites: MG12, MG09, MG07, MG05, MG04, and MG02. Among these sites, MG12 (Hafir Shehab El Din Pump Station) demonstrated the most substantial improvement, with its BWQI increasing from 46 (Poor) in 2000-2005 to 65 (Marginal) in 2017-2023, indicating the best water quality. MG09 also showed moderate improvement, with BWQI values fluctuating between25 (Poor) in 2000-2005 and 49 (Marginal) in 2017-2023. MG07 experienced considerable progress, with BWQI rising from 25 (Poor) to 46 (Marginal) by 2017-2023, reflecting a notable reduction in pollution.

In contrast, MG05 saw moderate improvement, with its BWQI increasing from 22 (Poor) in 2000-2005 to 35 (Poor) by 2017-2023. MG04 exhibited poor water quality throughout the study period, with BWQI values rising only slightly from 21 (Poor) to 29 (Poor). MG02 consistently had the worst water quality, with BWQI starting at 8 (Poor) and only reaching 23 (Poor) by 2017-2023.

This analysis indicates that the upstream branch drains contribute the most to pollution, with sites near the inlet of the drain (such as MG02, MG04, and MG05) consistently showing the poorest water quality. However, as the drain progresses downstream, water quality improves. Downstream branch drains, such as MG07, MG09, and MG12, contribute to this overall improvement. As a result, part of the water from these downstream branches (MG07, MG09, and MG12) may be suitable for reuse by mixing it with nearby canals before it flows into the main drain. Additionally, there has been a noticeable improvement in water quality across all sites over the years, reflecting ongoing efforts to reduce pollution.

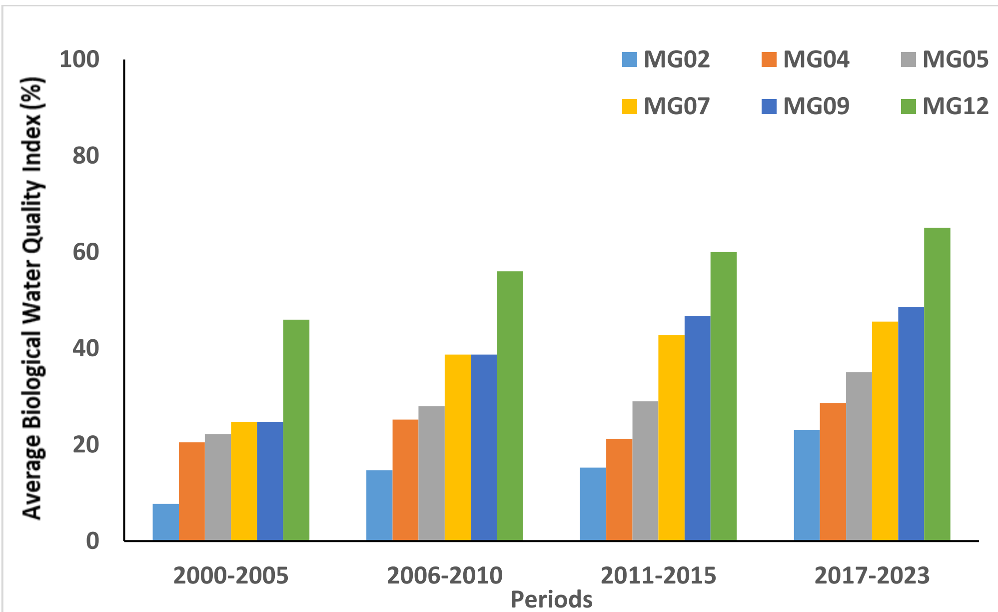


Figure 11: Average BWQI across Branch Sites (2000-2023)

3.7. Seasonal Variations in Biological Pollution Across Gharbia Main Drain Branches (2000-2023)

From 2000 to 2023, the seasonal variation in biological water quality across the branch sites of the Gharbia Main Drain, as presented in **Figure 12a–12f**, reveals distinct patterns. At MG02, summer and spring generally exhibited better biological water quality compared to fall and winter. Summer BWQI increased from 6 in 2000–2005 to 33 in 2017–2023, while spring BWQI fluctuated, starting at 11 in 2000–2005, then dropping to 10 in 2011–2015, before recovering to 24 in 2017–2023. Winter and fall consistently showed lower biological water quality with fall BWQI rising from 5 to 23, and winter BWQI showing minimal improvement, rising from 9 to 12 over the same period.

At MG04, summer and spring generally showed the highest BWQI values, reaching around 38 and 29 respectively by 2017–2023. However, winter values were significantly lower, ranging from 7 in 2000–2005 to 19.8 in 2017–2023, confirming that winter remains the poorest season at this site. Fall BWQI also remained relatively low, peaking at 29 in 2006–2010 before declining slightly to 28.1 in the most recent period. MG05 shows a more distinct seasonal variation, with summer and spring having better water quality, with summer BWQI improving from 25 to 47 and spring BWQI increasing from 18 to 40. In contrast, fall and winter showed lower BWQI values, with fall rising only slightly from 24 to 25, and winter BWQI increasing from 22 to 28.

For MG07, summer is the best season, with BWQI increasing from 39 to 50. Spring and fall also showed gradual improvement, with spring rising from 11 to 42 and fall improving from 29 to 45. Winter, while improving from 20 to 44.9, remained slightly lower than the other seasons, reflecting lower water quality. MG09 fall demonstrated the highest biological water quality in recent periods, with BWQI rising from 24 to 57 by 2017–2023. Spring also improved steadily, rising from 15 to 52. Summer and winter followed similar trends.

Finally, MG12 demonstrates the highest water quality results across all seasons, with significant improvements over time, particularly in the more recent periods, where most seasons reach much higher BWQI values, indicating excellent water quality. Spring consistently exhibited the highest BWQI values, improving from 51 in 2000–2005 to 92 in 2017–2023. Winter BWQI also improved substantially from 38 to 63 over the same period. Summer and fall showed less dramatic improvements, with summer stabilizing at around 54, and fall improving from 51 to 52. Among all the monitored sites, MG12 consistently demonstrated the best biological water quality across all seasons. To reinforce what has been mentioned in previous results, the water from MG09 and MG12, particularly during the spring season, can be reused by mixing it with nearby canals.

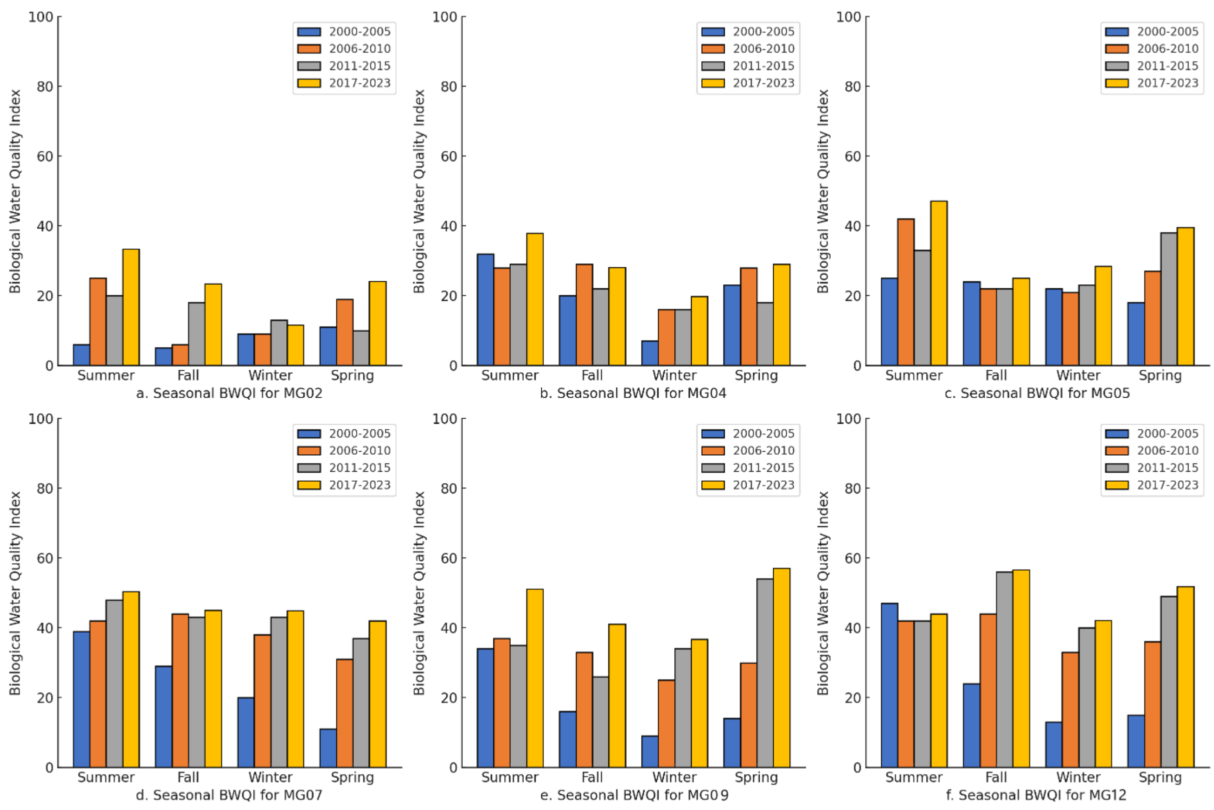


Figure (12a-12f) Seasonal Variation of BWQI across Branch Sites (2000-2023)

3.8. Feature Importance for Water Quality Parameters Results

As shown in **Table 7,** Fecal Coliform (FC) exhibited the highest feature importance for the biological category, contributing 61.56% to the overall pollution levels, followed by Biological Oxygen Demand (BOD) at 38.44%. This indicates that FC and BOD play a dominant role in the biological pollution load.

Table 7. Biological parameters independent variables importance

Data	Feature importance	Normalized importance
FC	61.56%	100%
BOD	38.44%	59.54%

For the industrial category, illustrated in **Table 8**, Chemical Oxygen Demand (COD) was identified as the most critical feature, contributing 80.01% to pollution levels. Other variables showed much smaller importance, highlighting the prominent role of COD in industrial waste affecting the water quality of the drain.

Table 8. Industrial parameters independent variables importance

Data	Feature importance	Normalized importance
COD	80.01%	100%
Cd	2.46%	3.07%
Cu	0.94%	1.17%
Fe	4.57%	5.71%
Mn	1.74%	2.17%
Zn	0.16%	0.199%
Ni	2.26%	2.82%
Pb	7.85 %	9.81%

In the case of the aquatic category, as placed in **Table 9**, Dissolved Oxygen (DO) emerged as the most influential factor, contributing a striking 96.19%. This dominance reflects the importance of oxygen depletion as a key indicator of aquatic pollution, while other variables, such as Total Suspended Solids (TSS) and pH, contributed less than 3%. Therefore, when modeling the aquatic category, DO alone is included.

Table 9. Aquatic parameters independent variables importance

Data	Feature importance	Normalized importance
DO	96.19%	100%
TSS	2.34%	2.43%
pH	1.47%	1.53%

Agricultural category, as indicated in **Table 10**, was primarily driven by Ammonium (NH_4), which had a feature importance of 30.72%, followed by Total Dissolved Solids (TDS) at 25.63% and Electrical Conductivity (EC) at 24.81%. Other variables, such as Total Nitrogen (TN) at 12.16%, Nitrate (NO_3) at 4.69%, and Total Phosphorus (TP) at 1.99%, contributed less significantly.

Table 10. Agricultural parameters independent variables importance

Data	Feature importance	Normalized importance
EC	24.81%	80.7%
TDS	25.63%	83.5%
TP	1.99%	6.5%
NO_3	4.69%	1.44%
NH_4	30.72%	100%
TN	12.16%	39.5%

3.9. Sensitivity Analysis for Water Quality Parameters Results

For biological category, as shown in **Figure 13**, sensitivity analysis revealed that changes in Fecal Coliform (FC) and Biochemical Oxygen Demand (BOD) have a pronounced effect on pollution levels. A 10% increase in FC and BOD levels led to a significant decrease in the Water Quality Index (WQI), indicating that biological pollution is highly sensitive to these fluctuations. This underscores the need to focus on managing sources of biological contamination to improve water quality.

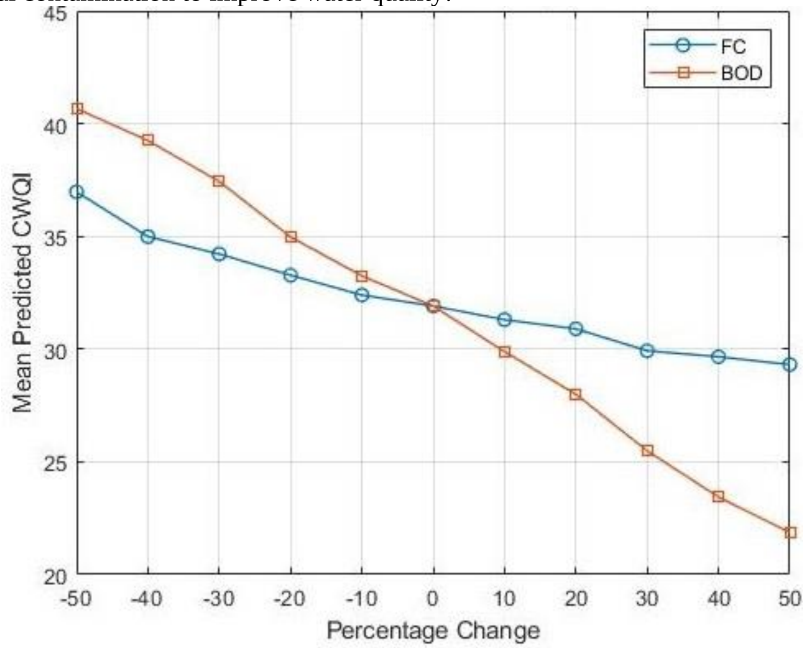


Figure 13: Biological Sensitivity Analysis

In the case of the industrial category, as presented in **Figure 14**, sensitivity analysis showed that Chemical Oxygen Demand (COD) is highly sensitive to changes. Even a small percentage change in COD levels resulted in noticeable shifts in the Water Quality Index (WQI), reaffirming COD's dominant role in affecting water quality. In contrast, other industrial parameters exhibited significantly lower sensitivity compared to COD.

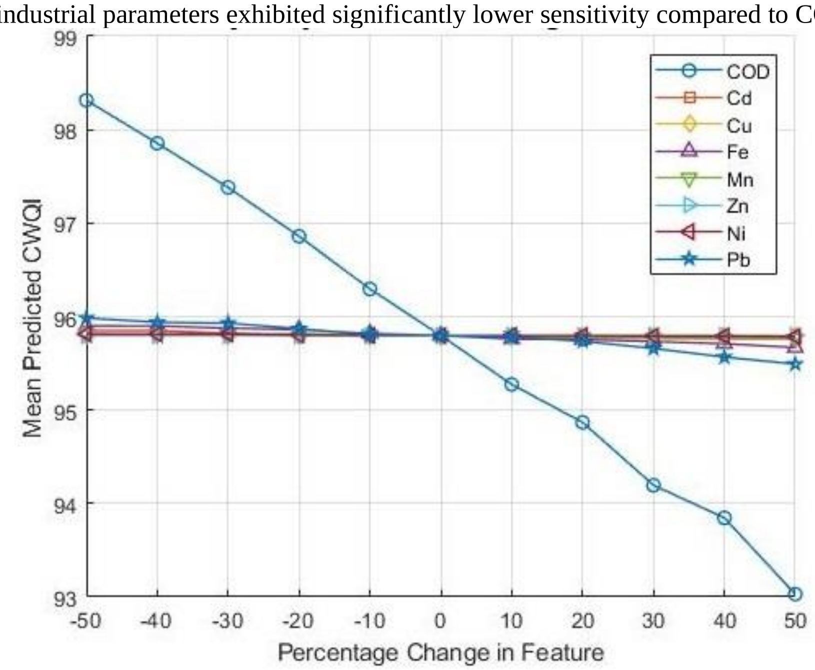


Figure 14: Industrial Sensitivity Analysis

For the aquatic category, as outlined in **Figure 15**, Dissolved Oxygen (DO) demonstrated extreme sensitivity to changes, with any variation in DO levels causing a substantial impact on the Water Quality Index (WQI). This suggests that improving biological pollution parameters is essential for maintaining adequate DO levels, which is crucial for supporting aquatic life in the drain.

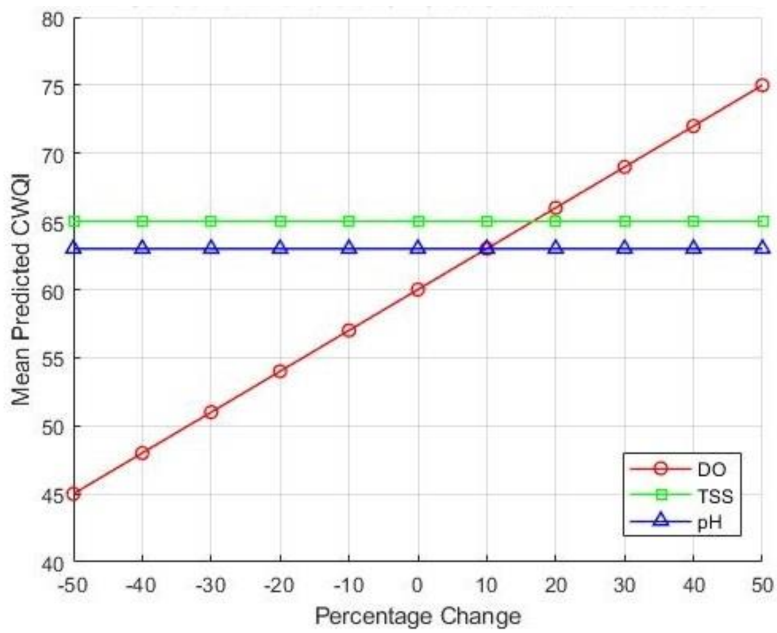


Figure 15: Aquatic Sensitivity Analysis

The agricultural category, as mentioned in **Figure 16**, including Ammonium (NH_4), Total Dissolved Solids (TDS), and Electrical Conductivity (EC), were also sensitive to changes. A 10% increase in these levels led to a marked rise in agricultural pollution, as measured by the Water Quality Index (WQI). The other parameters exhibited moderate sensitivity.

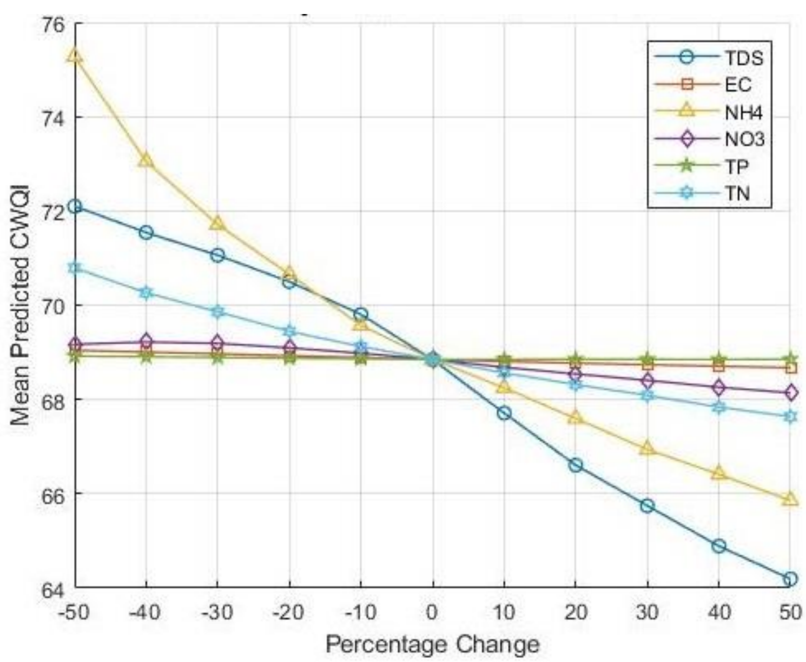


Figure 16: Agricultural Sensitivity Analysis

3.10. Comparison between Study Findings and Previous Research on Water Quality in Gharbia Main Drain

The findings of this study align with several key observations from previous research on the water quality of the Gharbia Main Drain while also presenting distinctive differences. Consistent with the studies conducted by (Osman et al., 2023),(Abd-Elfattah et al., 2021), and (Darwish et al. 2023), our research identified biological pollution, characterized by high Fecal Coliform (FC) and Biological Oxygen Demand (BOD) levels, as a dominant contributor to contamination in the drain. Despite agreeing with (El-Razak et al., 2023) that there are seasonal trends in pollution levels in the Gharbia drain, our results disagreed with this research in identifying the seasons during which the most severe water quality degradation occurs.

Additionally, while studies such as (El Gammal 2016), reported significant five water quality parameters as the most influential when evaluating Gharbia Drain water quality, our study identified seven key parameters. Furthermore, unlike previous studies that emphasized the unsuitability of the water for irrigation due to high salinity and microbial contamination (Darwish et al., 2023; El-Razik et al., 2023) along the whole Gharbia Drain, our findings indicate gradual improvements in water quality at downstream branches, suggesting the potential for reuse under specific conditions. Moreover, our study extends beyond past works by employing Artificial Neural Networks (ANNs) to predict Water Quality Indices (WQIs), demonstrating superior accuracy and predictive reliability compared to traditional statistical methods. These advancements highlight the efficacy of ANNs in capturing nonlinear relationships in water quality data, offering a more robust framework for water management decisions.

Conclusion

From the above research, the following can be concluded:

- The results of the feature importance and sensitivity analysis identified seven key parameters as the most influential when evaluating WQIs: Fecal Coliform (FC) and Biological Oxygen Demand (BOD) are the primary factors influencing biological pollution. In the industrial category, Chemical Oxygen Demand (COD) plays a dominant role. For aquatic pollution, Dissolved Oxygen (DO) is the most critical parameter, while Ammonium (NH_4), Total Dissolved Solids (TDS), and Electrical Conductivity (EC) are the key drivers of agricultural pollution.
- The most significant source of pollution in the Gharbia Drain is driven by biological pollutants, while agricultural pollution is moderate. On the other hand, there is no significant industrial pollution observed. Additionally, across all studied seasons, summer has the best water quality, followed by spring. Winter shows moderate improvement but still has higher pollution levels. Fall is the worst season.
- The research investigated water quality parameters over different periods (2000-2005), (2006-2010), (2011-2015), and (2017-2023). In recent years, a steady reduction has been observed in biological pollution over time. This aligns with the government's efforts to establish wastewater treatment plants in the Gharbia Main Drain catchment area.
- Segaaya P.S., Samatay P.S., and P.S. No. 5 Branch drains (located at the upstream reach of the drain) are the major contributors to pollution, showing the poorest water quality indices. In contrast, downstream branch drains (P.S. No. 6, P.S. No. 4, and Hafir Shehab El Din P.S.) exhibit better water quality, making them suitable for water reuse by mixing it with nearby canals before entering the main drain.

Recommendations for Further Studies

Based on the study's conclusion that biological pollution is a major concern, the following recommendations are provided for wastewater treatment and policy interventions to improve water quality in the studied drain. A multifaceted approach to wastewater treatment is essential; establishing a thorough wastewater treatment system can effectively eliminate biological contaminants from the water. This approach should encompass primary, secondary, and tertiary treatment phases to ensure compliance with established quality standards. Additionally, the implementation of policies that encourage responsible waste disposal is crucial. The decision tree methodology, a well-established technique in decision-making, involves constructing a tree-like diagram that outlines decisions and their potential outcomes. This methodology is considered instrumental in supporting decisions that improve water quality in drainage systems and mitigate the sources of biological pollution.

List of Abbreviations

Abbreviation	Water Quality Parameter
BOD	Biological Oxygen Demand
Cd	Cadmium
COD	Chemical Oxygen Demand
Cu	Copper
DO	Dissolved Oxygen
EC	Electrical Conductivity
FC	Fecal Coliform
Fe	Iron
Mn	Manganese
NH_4	Ammonium
Ni	Nickel
NO_3	Nitrate
Pb	Lead
pH	Hydrogen Ion Concentration
TDS	Total Dissolved Solids
TN	Total Nitrogen
TP	Total Phosphorus
TSS	Total Suspended Solids
Zn	Zinc

References

Abd-Elfattah EA, Sheta AEA, Saifeldeen M, Hassanein SA, Mahmoud YI. 2021. “Assessment of Water and Sediments Quality of Kitchener Drain, Nile Delta, Egypt.” *Arab Universities Journal of Agricultural Sciences* 29(2): 801–11. doi:10.21608/ajs.2021.55062.1320.

Abdel-Fattah MK, Mokhtar A, Abdo AI. 2021. “Application of Neural Network and Time Series Modeling to Study the Suitability of Drain Water Quality for Irrigation: A Case Study from Egypt.” *Environmental Science and Pollution Research* 28: 898–914. doi:10.1007/s11356-020-10543-3.

Abdelrazek S. 2019. “Monitoring Irrigation Water Pollution of Nile Delta of Egypt with Heavy Metals.” *Alexandria Science Exchange Journal* 40(July-September): 441–50. doi:10.21608/asejaiqjsae.2019.50350.

Abosena A, Abbas H, Farid I, El-Kholy M. 2021. “Environmental Assessment of El-Gharbia Main Drain Water.” *Environment, Biodiversity and Soil Security* 5: 185–203.

Allam A, Tawfik A, Yoshimura C, Fleifle A. 2016. “Multi-Objective Models of Waste Load Allocation toward a Sustainable Reuse of Drainage Water in Irrigation.” *Environmental Science and Pollution Research* 23(12): 11823–34. doi:10.1007/s11356-016-6331-z.

Altunkaynak A. 2007. “Forecasting Surface Water Level Fluctuations of Lake van by Artificial Neural Networks.” *Water Resources Management* 21(2): 399–408. doi:10.1007/s11269-006-9022-6.

Amin MA. 2002. "Predicting the Variations in Water Quality along an Irrigation Canal in Punjab, Pakistan." McGiH University.

Canadian Council of Ministers of the Environment. 2001. "Canadian Water Quality Guidelines for the Protection of Aquatic Life: CCME Water Quality Index 1.0, Technical Report." *Canadian Council of Ministers of the Environment*: 1–13. https://unstats.un.org/unsd/envaccounting/ceea/archive/Water/CCME_Canada.PDF.

Darwish, Dina H., Mahy M. Ameen, Abeer M. Salama, Mokhtar S. Beheary, and Mamdouh S. Serag. 2023. "Ecological Risk Assessment of Heavy Metals in Water, Sediment and Macrophytes of Two Drains in the Deltaic Mediterranean Coast of Egypt." *Egyptian Journal of Aquatic Biology and Fisheries* 27(5): 137–70. doi:10.21608/ejabf.2023.317457.

Dehkordi DK, Kashkuli HA. 2015. "Water Quality Improvement of Hendijan River by Use of Mike11 Model." *Advances in Environmental Biology* 9(3): 319–27.

Donia N, El Gammal H, Abdelkawi K, El-Bahrawy A. 2009. "EVALUATION OF THE REUSE OF DRAINAGE WATER USING MODELING TECHNIQUES CASE STUDY: GHARBIA DRAIN, EGYPT." *Thirteenth International Water Technology Conference*: 1–8. doi:http://water-observatory.net/sources/iwtc2009/19-2.PDF.

Drainage Research Institute (DRI) year books 2000:2015. "No Title." *Drainage water status in the nile delta : yearbook 2000/ 2015*.

Egypt Decree. 2013. *Egypt Decree, 92/2013. For The Protection of the Nile River and Its Waterways from Pollution. Decree of Minister of Water Resources and Irrigation No. 92 for Year 2013 for the Executive Regulation of Law 48/1982*.

EGYPT Decree Law 48/. 1982.

El-Amier YA, Kotb WK, Bonanomi G, Fakhry H, Marraiki NA, Abd-Elgawad AM. 2021. "Hydrochemical Assessment of the Irrigation Water Quality of the El-Salam Canal, Egypt." *Water (Switzerland)* 13(17): 2428. doi:10.3390/w13172428.

El-amier, Yasser A, Mahmoud A Zahran, Ahmed S Gebreil, and Eman H Abd El-salam. 2017. "Anthropogenic Activities and Their Impact on the Environmental Status of Anthropogenic Activities and Their Impact on the Environmental Status of Kitchener Drain , Nile Delta , Egypt." *J. Environ. Sci* 46: 251–62.

El-Gammal, H., Hatem, A., & EL-Bahrawy, A. 2009. "Assessment of Water Reuse As a Non-Conventional Water Resource in Egypt Case Study: Gharbia Drain, Nile Delta." In *Thirteenth International Water Technology Conference, IWTC13 2009, Hurghada, Egypt*, http://water-observatory.net/sources/iwtc2009/19-2.PDF.

El-Razik AM, Amin OF, Ramadan MSA, Abosena AM. 2023. "A Qualitative and Quantitative Study to Effect of El-Gharbia Main Drain Wastewater on The Surrounding Soils and Plant Life." *Journal of Soil Sciences and Agricultural Engineering* 14(4): 133–41.

El-razik, Eman M Abd, Omnia F Amin, M S A Ramadan, A M Abosena, and Cross Mark. 2023. "Journal of Soil Sciences and Agricultural Engineering A Qualitative and Quantitative Study to Effect of El- Gharbia Main Drain Wastewater on The Surrounding Soils and Plant Life." 14(4): 133–41. doi:10.21608/jssae.2023.196540.1151.

El-Sayed A, Shaban M. 2019. "Developing Egyptian Water Quality Index for Drainage Water Reuse in Agriculture." *Water Environment Research* 91(5): 428–40. doi:10.1002/wer.1038.

El-Sherbiny EK, El-Kassas H, El-Saadi AM. 2018. "Evaluation of El-Gharbia Drain Water Quality To Increase Benefits From It." *Journal of Environmental Science* 43(1): 71–104. doi:10.21608/jes.2018.30846.

Esraa, Elsadek, Noureldin Mohamed, Shaltout Fatma, and Balah Ahmed. 2023. “WEAP Analysis for Enhancing Water Resource Sustainability in Egypt: A Dynamic Modeling Approach for Long-Term Planning and Management.” *HBRC Journal* 19(1): 253–74. doi:10.1080/16874048.2023.2260602.

El Gammal, H A A. 2016. “Statistical Analysis of Water Quality Monitoring Network Case Study: Gharbia Drainage Catchments Area.” *Advances in Environmental Biology* 10(4): 297–305.

Gavali KR, Gundale AS. 2023. “Water Quality Evaluation Using Machine Learning Techniques.” : 1–10.

Hamed MA. 2019. “Characterization of Surface Water Quality along Ismailia Canal, Nile River, Egypt.” *Journal of Advanced Civil and Environmental Engineering* 2(1): 01. doi:10.30659/jacee.2.1.01-14.

Ibrahim J, Chen MH, Sinha D. 2009. 27 The Elements of Statistical Learning *Springer Series in Statistics*. http://www.springerlink.com/index/D7X7KX6772HQ2135.pdf.

Juahir H, Zain SM, Toriman ME, Mokhtar M, Man HC. 2004. “Application of Artificial Neural Network Models for Predicting Water Quality Index.” *Jurnal Kejuruteraan Awam* 16(2): 42–55.

K., Mosimanegape. 2016. “Integration of Physicochemical Assessment of Water Quality with Remote Sensing Techniques for the Dikgathong Dam in Botswana.”

Khadr M, Elshemy M. 2017. “Data-Driven Modeling for Water Quality Prediction Case Study: The Drains System Associated with Manzala Lake, Egypt.” *Ain Shams Engineering Journal* 8(4): 549–57. doi:10.1016/j.asej.2016.08.004.

Khalifa AK, Abdel H, El H. 2006. “Evaluation of the Reuse of Drainage Water Using Modeling Techniques Case Study : Gharbia Drain , Egypt.” *Environmental Studies*: 1–8.

Kişi Ö. 2006. “Evapotranspiration Estimation Using Feed-Forward Neural Networks.” *Nordic Hydrology* 37(3): 247–60. doi:10.2166/nh.2006.010.

Konaté AA, Pan H, Khan N, Yang JH. 2015. “Generalized Regression and Feed-Forward Back Propagation Neural Networks in Modelling Porosity from Geophysical Well Logs.” *Journal of Petroleum Exploration and Production Technology* 5(2): 157–66. doi:10.1007/s13202-014-0137-7.

L., Breiman. 2001. “Random Forests.” : 5–32.

Mbeche GO. 2021. “WATER QUALITY ASSESSMENT OF NYAKOMISARO TRIBUTARY OF RIVER KUJA, KISII COUNTY, KENYA.” Kenyatta University.

MM., Sahoo. 2014. “Analysis and Modelling of Surface Water in River Basins (Doctoral Dissertation).”

Mohamed, M, A Elansary, and M Moussa. 2017. “A Modelling Approach to Manage Water Quality at Gharbia Main Drain, Egypt.” In *Twentieth International Water Technology Conference, IWTC20, Hurghada, Egypt*, , 133–45.

NWQM. 2003. “National Water Quality & Availability Management Project.” *Website: www.nwrc-egypt.org* 18(10): 97–98, 102, 104.

Osman WS, Ali D, Bhran AA, Hassanean MH, Shoaib AM. 2023. “Management of Physical and Chemical Water Quality in Kotchener Drain (Kafr El Sheikh, Egypt).” *Asia-Pacific Journal of Chemical Engineering* 18(5): e2964.

Radwan E, Eissa E, Nassar AM, Salim Y, Hashem H, Abdul-Aziz K, Abdel-Hakeem N. 2019. “Study of Water Pollutants in El-Mahmoudia Agricultural Irrigation Stream at El-Beheira Governorate, Egypt.” *Journal of Bioinformatics and Systems Biology* 02(01): 1–18. doi:10.26502/jbsb.5107004.

S., Awang. 2015. “A Water Quality Study of the Selangor River, Malaysia.” University of East Anglia.

Singh KP, Basant A, Malik A, Jain G. 2009. “Artificial Neural Network Modeling of the River Water Quality-A Case Study.” *Ecological Modelling* 220(6): 888–95. doi:10.1016/j.ecolmodel.2009.01.004.

Taha AA, El-Shehawy ME, Mosa AA, El-Komy MN. 2012. “Suitability of Drainage Water for Irrigation and Its Impact on Wheat and Clover Crops At Northern Delta, Egypt.” *Journal of Soil Sciences and Agricultural Engineering* 3(6): 655–68. doi:10.21608/jssae.2012.54343.

Thakur D, Singh A. 2023. “Exploring Machine Learning Algorithms for Reliable Water Quality Prediction.” *International Journal for Research in Applied Science and Engineering Technology* 11(September): 1437–43.

Tiyasha, Tran MT, Yaseen ZM. 2020. “A Survey on River Water Quality Modelling Using Artificial Intelligence Models: 2000–2020.” *Journal of Hydrology* 585: 124670. doi:10.1016/j.jhydrol.2020.124670.

Tyagi S, Sharma B, Singh P, Dobhal R. 2013. “Water Quality Assessment in Terms of Water Quality Index.” *American Journal of Water Resources* 1(3): 34–38. doi:10.12691/ajwr-1-3-3.

Vardhan KH, Kumar PS, Panda RC. 2019. “A Review on Heavy Metal Pollution, Toxicity and Remedial Measures: Current Trends and Future Perspectives.” *Journal of Molecular Liquids* 290: 111197. doi:10.1016/j.molliq.2019.111197.

W.O, Nassar, Rasha Hosny, Mohammed Ghareeb, Hany F. Abd-Elhamid, Martina Zeleňáková, and Manal Gad. 2025. “Assessment of Groundwater Quality and Contamination Hazardous Using Water Quality Index and Physicochemical Analysis in Assiut Governorate, Upper Egypt.” *Water, Air, and Soil Pollution* 236(3). doi:10.1007/s11270-025-07789-3.

Zaghloul SS, Elwan H. 2011. “Water Quality Deterioration of Middle Nile Delta Due to Urbanizations Expansion, Egypt.” *Fifteenth International Water Technology Conference, IWTC-15 2011*: 17. http://iwtc.info/wp-content/uploads/2011/07/G40.pdf.

Zhang L, Pang J, Wang Y, Yang M, Shi Y, Yang L. 2010. “SPSS for Water Quality Assessment of Beijing Typical River Based on Principal Component Analysis.” In *2010 International Conference on Digital Manufacturing & Automation*, IEEE, 395–98. doi:10.1109/ICDMA.2010.377.