Machine learning analysis for the Hilla River water quality index - Iraq

Hussien A. M. AL-Zubaidi[1], Entidhar Jawad Kadhim[2], Hadeel Kareem Jasim[3], Ahmed Samir Naje[*4], Fatimah D. Al-Jassani[5], Shreeshivadasan Chelliapan[6]

[1]Department of Environmental Engineering, Faculty of Engineering, University of Babylon, Iraq

[2]Civil Engineering Department, College of Engineering, Al-Qasim Green University, Babylon 51013, Iraq

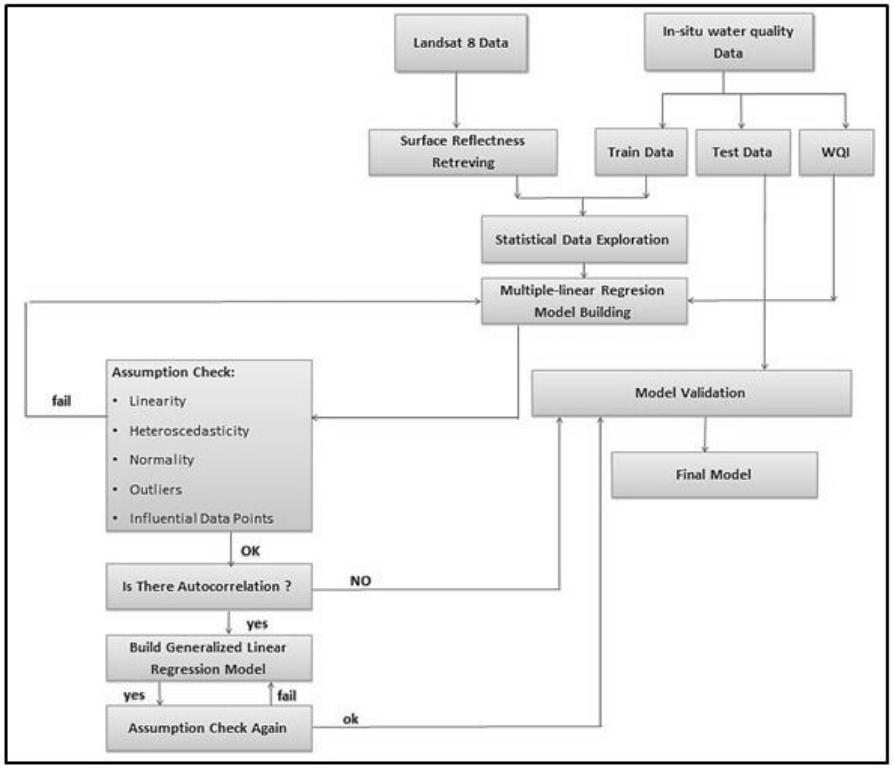
[3,4]Water Resources Management Engineering Department, College of Engineering, Al-Qasim Green University, Babylon 51013, Iraq

[5]Department of Environmental Engineering, Faculty of Engineering, University of Babylon, Iraq

[6]Department of Engineering, UTM Razak School of Engineering and Advanced Technology, Universiti Teknologi Malaysia, Jalan Semarak, 54100, Kuala Lumpur, Malaysia

*Corresponding Author: ahmednamesamir@yahoo.com

Graphical Abstract



Abstract

Since industrial and human activities have been developed in different ways in Iraq, water quality has been declining along the Hilla River, the only water resource for drinking water in the Hilla City, Iraq. In this research, the Weighted Arithmetic Water Quality Index (WQI) along the river was analyzed using the linear regression machine learning algorithm. Water quality parameters including Turbidity (Turb), Electric Conductivity (EC), Hydrogen Ions (pH), Total Suspended Solid (TSS), Chloride Ions (Cl), Sulfate Ions (SO4), Alkaline (ALK), Total Hardness (TH), Calcium Ions (Ca), Potassium Ions (k), Sodium Ions (Na), Magnesium Ions (Mg), and Total Dissolved Solid (TDS) were utilized to determine WQI from January 2016 to June 2021 depending on datasets from five sampling stations located along the river at the Hilla City. It was noticed that the river WQI has a significant relationship with Turb only (positive proportion). This relationship between WQI and Turbidity in the river is limited to aWQI value of 220, Thus, two linear regression models were developed and validated: One for WQI values greater than 220 and another for the values less than 220. In addition, the results of this study showed that the Hilla River is severely polluted since its WQI values are high. The best WQI and turbidity value were in 2018. However, in 2020 and 2021, there were some improvement in WQI and turbidity compared to 2019. Hence, splitting the water quality dataset provides a way to find a significant correlation with WQI.

Keywords: Correlation, Linear regression, Machine learning, Water quality, Water quality index

1. Introduction

The most valuable resource that nature offers to humans is water. It is crucial to our survival and necessary for all human endeavors, including agriculture, trade, industry, the production of power, daily hydration, and other activities (Abbas, 2013; Kannan & Ramasubramanian, 2011). In Iraq, water bodies cover more than 5% of the country's surface (Abbas et al., 2018), including lakes, reservoirs, and rivers (Tigris, Euphrates, and Shatt al-Arab), as well as regions with stagnant water such as small lakes and marshes. However, over the past two or three decades, as human activity has expanded, the quality of water in numerous large rivers has decreased internationally (Avid Hirst & ob Morris, 2001; Ochir & Davaa, 2011). River water quality declines as a result of numerous factors including unidentified causes (Carpenter et al., 1998; Kotti et al., 2005). Accordingly several physical, chemical, and biological indicators can describe the type and extent of water pollution (Chitmanat & Traichaiyaporn, 2010). One of the most important techniques for categorizing and distributing water quality data to the general public and the appropriate decision-makers is the water quality index (WQI) (Mohammed & Shakir, 2012; Oko et al., 2014; Shanmuga Sundaram, et al., 2024).

Al-Ridah et al. (2020) used the Water Quality Index of the Canadian Council of Ministers of the Environment and the Weighted Arithmetic models to assess the water quality for drinking (CCME WQI). The outcomes of these two models were compared as well. Four water treatment

facilities on the Hilla River, a tributary of the Euphrates River in central Iraq, were included in the study area. From January to December 2018, water samples were taken on a monthly basis, and nine parameters of raw water were examined, such as turbidity (Tur), pH, electric conductivity (EC), alkalinity (Alk), total hardness (TH), calcium (Ca), magnesium (Mg), chloride (Cl-), and total dissolved solids (TDS). For all stations, the Weighted Arithmetic model showed that the raw water quality was categorized from "severely polluted" to "unfit for human consumption". However, the CCME WQI method categorized the river water as "fair" and treated water as "good" for drinking. The comparison results of two models showed that CCME WQI gave greater water quality value than the value from the other method, or the CCME WQI was possibly considered as more flexible. Al-Bayati et al. (2018) investigated field spector-radiometers by developing relationships between water quality parameters and spectral data. The study included 20 stations for sampling on Hilla River, Babylon Province, Iraq to measure the physical and chemical parameters (pH, TSS, EC, TDS, and CL). Landsat 8 satellite images were employed to be linked with field data statistically for only one day of investigation. It has been found that apposite spectral ranges and bands for water quality parameters, EC and CL associated with a spectra range of (0.851–0.87) μm and (2.107–2.294) μm, respectively. Also, (TSS and Turb), and TDS at a spectral range of (0.533–0.590) μm and (1.566–1.561) μm, respectively. Furthermore, Chabuk et al. (2020) assessed the Tigris River's WQ utilizing the WQI method. Samples of 12 variables were collected from 14 sampling stations along the river. The water quality index was calculated using the weighted arithmetic approach (WQI). On three locations along the Tigris River in both seasons, the regression prediction was utilized to compare actual values with those predicted by the prediction maps. The findings showed that all parameters' regression forecasts got sufficient determination coefficient values (R^2). Furthermore, during both winter and summer seasons, the WQ for the Tigris River deteriorated following of the River flow, especially at the station (8) in Aziziyah, with obvious increases in degradation at Qurnah (Basrah province) in southern Iraq. The full length of the Tigris River is considered for the purposes of this study. This is critical for providing full understanding about the river's pollution reality. As a result, it is easier to recognize the contamination problem, assess it, and then find appropriate treatments and solutions.

Recently, machine learning has been utilized widely to predict water quality indices or parameters based on many water quality features. Latest researches have shown that machine learning approaches with their substantial ability for identifying the important features are being broadly used for the water quality prediction (Venkatraman et al., 2024; Sundarapandi et al., 2024; Jegan et al., 2024; Babu et al., 2024; Venkatraman and Surendran, 2023). Therefore, developing a WQI efficient linkage technique with the water quality parameters is very important for water quality monitoring. It transforms complex data of water quality into information that the public can understand and use. Thus, the main objectives of this research are to analyze the Hilla River water quality based on its water quality parameters that have a significant relationship with WQI by implementing the linear regression machine learning technique uniquely.

2. Materials and methods

2.1. Study area and datasets

The city of Hilla is located in the center of Iraq where the ancient city of Babylon is located in Babylon Province, Iraq. It is situated in a predominantly agricultural area that receives extensive irrigation from the Hilla River (Al-Ridah et al., 2021; Al-Saadi and Al-Zubaidi, 2024), producing a broad variety of vegetables, fruit, and textiles. Hilla River is considered a branch from the Euphrates River at Saddat Al-Hindiyah Reservoir (Al-Dalimy and Al-Zubaidi, 2023). Figure 1 shows the present study area. It is situated between, Longitude (44º26º55" & 44º31º10") E and Latitude (32º26º30" & 32º31º33") N. Table 1 depicts the selected sampling station points along the river (S1, S2, S3, S4, and S5). Samples were collected sparsely by Babylon Water Resources Directorate, Iraq at each sampling stations from 2016 to 2021. The collection process has included taking one or two samples monthly during this period. Following the standard methods of water examination (APHA, 2017), the following devices were used in lab to test the water samples: pH meter, EC meter, Turb meter; Spectrophotometer, Flame photometer, Burettes, Electric balance, Elictric drying oven) in addition to the titration device. Table 3 to 6 review the stations water quality parameters yearly average values during the study period.

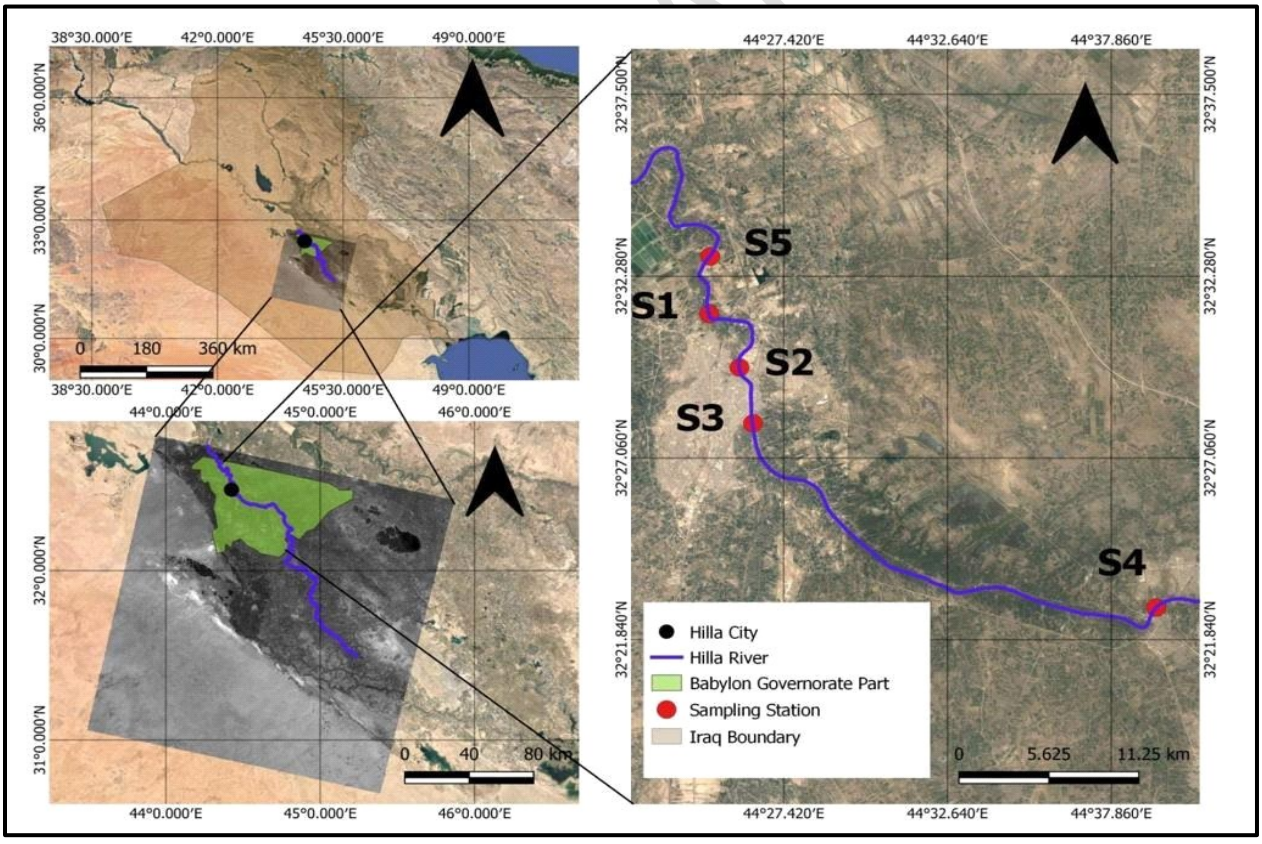


Figure 1. Study area and sampling location.

Table 1. Sampling station along Hilla River.

No	Station	Location
1	S1	The New Hilla Water Treatment Plant
2	S2	The Old Al-Tayarah Water Treatment Plant
3	S3	Al-Hashmiya Water Treatment Plant
4	S4	Al-Atayej Water Project
5	S5	The Annanah Water Project

Table 2. The yearly average value of water quality parameters for S1 location.

Year	Parameter												
	Cl	ALK	TSS	k	Na	TDS	SO4	Mg	Ca	TH	EC	Turb	pH
2016	124.6	119.4	49.1	3.6	81.5	685.1	291.7	39.8	86.8	379.6	1063.7	13.8	7.8
2017	104.7	126.6	32.4	3.0	79.3	653.4	249.5	31.5	79.2	331.6	1015.5	7.4	7.7
2018	124.6	133.3	36.9	3.3	89.7	691.7	265.2	35.6	85.6	360.1	1142.7	5.9	7.1
2019	91.6	136.4	49.8	3.2	68.4	643.4	267.9	30.3	99.5	372.9	1020.8	28.9	7.3
2020	83.1	120.4	46.4	3.3	66.2	593.0	251.4	34.4	95.4	372.4	950.4	13.1	7.6
2021	88.8	104.8	26.9	3.4	72.2	611.0	259.4	35.1	85.3	356.6	997.9	7.4	7.9

Table 3. The yearly average value of water quality parameters for S2 location.

Year	Parameter												
	Cl	ALK	TSS	k	Na	TDS	SO4	Mg	Ca	TH	EC	Turb	pH
2016	122.8	119.5	50.0	3.2	80.7	675.5	288.1	40.0	86.5	377.7	1061.1	16.2	7.8
2017	104.8	129.3	33.1	3.0	79.8	634.1	245.7	31.4	77.8	328.5	1013.0	12	7.7
2018	122.8	133.7	44.9	3.3	90.0	682.3	266.2	35.5	84.7	356.8	1131.3	8	7.2
2019	94.7	133.5	48.6	3.2	70.0	647.9	271.6	30.8	99.5	375.0	1030.0	20.3	7.3
2020	85.1	119.2	61.2	3.3	69.7	607.5	261.0	33.1	96.4	377.0	967.1	13.7	7.4
2021	88.9	104.4	51.4	3.4	73.2	622.1	269.1	35.3	86.0	359.5	1004.1	15.3	7.9

Table 4. The yearly average value of water quality parameters for S3 location.

Year	Parameter												
	Cl	ALK	TSS	k	Na	TDS	SO4	Mg	Ca	TH	EC	Turb	pH
2016	125.6	118.1	61.0	3.4	83.7	675.0	294.9	38.4	88.2	376.1	1069.2	18.1	8.0
2017	104.3	126.8	37.7	2.9	77.8	636.3	244.8	32.9	77.2	332.8	997.0	10.8	7.6
2018	125.6	130.9	36.6	3.4	87.3	695.0	269.8	36.4	88.1	367.7	1154.6	5.6	6.9
2019	89.1	134.5	41.8	3.1	66.2	631.0	274.2	31.1	101.3	378.9	1017.8	15.8	7.3

2020	85.1	119.2	61.2	3.3	69.7	607.5	261.0	33.1	96.4	377.0	967.1	13.5	7.4
2021	87.8	104.8	28.8	3.4	70.3	608.0	258.0	33.3	84.6	347.6	991.0	8.5	7.9

Table 5. The yearly average value of water quality parameters for S4 location.

Year	Parameter												
	Cl	ALK	TSS	k	Na	TDS	SO4	Mg	Ca	TH	EC	Turb	pH
2016	128.3	118.9	53.7	3.7	85.9	691.2	306.4	40.3	90.7	392.4	1103.8	15.6	7.8
2017	107.1	128.1	36.8	3.0	81.5	646.3	254.3	32.8	78.9	335.7	1024.9	7.3	7.7
2018	128.3	135.5	43.7	3.6	88.3	701.3	259.5	35.3	85.7	360.0	1149.9	5	7.1
2019	100.8	133.5	40.9	3.8	73.7	625.5	246.5	31.6	99.6	373.6	1041.4	16.7	7.3
2020	88.3	126.7	43.3	3.7	72.3	586.7	223.3	37.7	83.0	362.3	931.0	16.3	7.7
2021	93.8	112.8	23.7	3.5	78.0	626.2	253.0	37.0	81.3	361.2	1013.0	8.78	7.6

Table 6. The yearly average value of water quality parameters for S5 location.

Year	Parameter												
	Cl	ALK	TSS	k	Na	TDS	SO4	Mg	Ca	TH	EC	Turb	pH
2016	128.2	120.6	51.4	3.3	83.5	668.8	290.1	38.8	86.3	372.9	1051.1	13.1	8.0
2017	108.4	128.1	35.4	3.0	82.4	628.8	249.1	32.2	79.9	336.3	1017.4	7.6	7.8
2018	128.2	132.0	40.2	3.3	90.7	698.7	269.5	36.7	87.0	366.6	1157.5	5	7.1
2019	93.9	134.5	34.0	3.1	71.7	654.5	268.2	31.7	98.1	374.6	1031.5	16.4	7.4
2020	83.3	135.5	32.0	3.6	67.3	554.0	218.3	36.0	82.8	353.8	910.3	13.3	7.8
2021	89.1	104.8	35.7	3.3	73.8	637.5	271.4	33.5	89.9	362.0	1018.1	10.4	7.9

2.2. The machine learning algorithm outline

The general outline of the machine learning algorithm used in this study was summarized in Figure 2. The WQI was calculated from in-situ water quality dataset for the five sampling stations on Hilla River. Then, the water quality parameters and WQIs were split into two sets: Train and test dataset. The linear regression model was trained between the WQI values and the other water quality parameters by using the R software statistical packages in order to find the best significant relationship (Al-Zubaidi et al., 2021). The pseudocode of linear regression model development process is presented in Algorithm I.

Algorithm I: The linear regression model development steps:

- **Start**
  - ***Reading*** *the water quality measurements dataset*
  - ***Exploring*** *and* ***cleaning*** *the dataset and* ***creating*** *a graphical summery*

- ***Calculating*** *the WQI (Eq. 1) and* ***appending*** *it as a feature to the dataset*
- ***Splitting*** *the entire dataset into: Train and Test dataset*
- ***Selecting*** *the water quality parameters that have a significant relationship with WQI using the Train dataset to be used to develop a multiple linear regression model*
- ***Validating*** *the developed model by using the Test dataset based on error statistics*
- ***Making*** *predictions depending on the final validated model*

- **End**

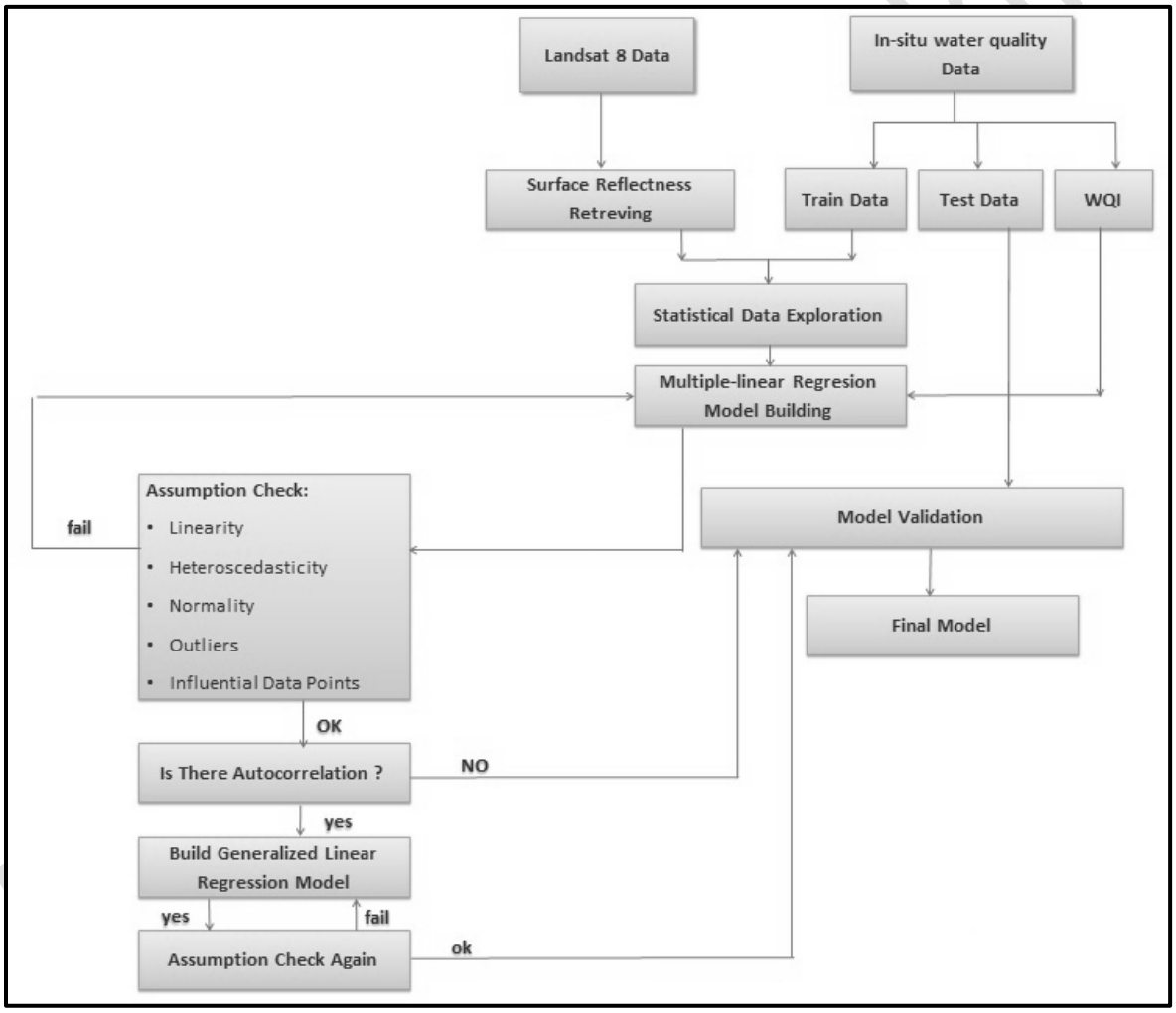


Figure 2. Data processing flowchart.

2.3. Water Quality Index (WQI) calculation

The WQI was introduced and defined as a calculated form of choosing, ranking, and mixing the important physical, chemical, and biological factors of water in a simple way in the mid-

twentieth century (Chabuk et al., 2022). Water quality is an important criterion in matching demand and supply water, and give expression simpler and easier to interpret data observations. Several Water Quality Indicators were utilized to evaluate the quality of surface water. However, the well-known one is the Arithmetic Weighted Water Quality Index (WQI). This method categorized the water quality according to the degree of pollution by using the most normally measured water quality and has been widely used by the various scientists (Pathak et al., 2015). Hence, the calculation of water quality index (WQI) was performed by using the following equation (Alobaidy et al., 2010; Kankal et al., 2012; Shanmuga Sundaram et al., 2024).

$$WQI = \frac{\sum_{i=1}^{n} W_i q_i}{\sum_{i=1}^{n} W_i} \quad (1)$$

Where n is the total number of water quality parameters, w_i is the relative weight of the i^{th} parameter, q_i is WQ rating of the i^{th} parameter, and W_i is the unit weight of WQ parameter.

3. Results and Discussion

3.1. In-situ measurements correlation and exploration

The entire water quality parameters measured during the study period were explored and depicted as shown in Figure 3, revealing the distribution of all parameters in addition to the correlation among them. Some of water quality parameters show a significant correlation with each other. Electrical conductivity (EC) has a significant correlation with three water quality parameters (chloride ions (Cl), sulfate ions (SO4) and total dissolved solid (TDS))with a value of r more than 70% and p-value less than 0.05. Total hardness (TH) shows a significant correlation with three water quality parameters (calcium ions (Ca), total dissolved solid (TDS) and sulfate ions (SO4)(with avalue of r more than 70% and p-value less than 0.05. Total dissolved solid (TDS) is linked with a significant correlation with four water quality parameters (chloride ions (Cl), sulfate ions (SO4), total hardness (TH) and electric conductivity (EC) (with a value of r more than 70% and p-value less than 0.05.

Accordingly, for the station S1, the maximum average yearly values of pH were in 2021 since these values were higher during dry seasons. The maximum value of Turb concentration was in 2019, where the Turb value variation in the same year was varying from (5.8 to 20.3) mg/L. The high value of EC was recorded in 2018; however, EC value variation through the study period was very small. The values of the TH, Mg, Ca, SO4, TDS, Na, K, and ALK are approximately remain constant or with very small change during the years of study. The maximum value of TSS and Cl was in 2019, and the two values have the high concentration in the wet season as shown in Table 2. For station S2, the maximum average yearly values of pH were also in 2021. The maximum value of Turb concentration was in 2019 in which the Turb value variation in the same year ranged from 8.0 to 52.6 mg/L. The high value of EC was recorded in 2018, but EC value variation through the study period was very small. The values of the TH, Mg, Ca, SO4, TDS, Na, K, and ALK were approximately constant or have a small change during the years of study. The maximum

concentration value of TSS was in 2020, and Cl has the same maximum value in 2016 and 2018. For station S3, the highest average annual pH levels were also recorded in 2016. The highest Turb concentration recorded in 2019. Although the EC value reached a high point in 2016, there was relatively little change over the course of the study period. During the years of investigation, the values of the following parameters remained steady or barely changed (TH, Mg, Ca, SO4, TDS, Na, K, and ALK). TSS has the same maximum concentration value in 2016 and 2020 while Cl has the same maximum concentration value in 2016 and 2018. For station S4, the highest average annual pH readings occurred in 2016 during the study period. In 2019 and 2020, turb's maximum value is the same. Although the EC value reached a high point in 2018, there was relatively little change over the course of the study period. The values of TH, Mg, Ca, SO4, TDS, Na, K, and ALK have remained rather consistent over the years of investigation or have very slightly changed. Cl has the same maximum value in 2016 and 2018 as TSS, which had the highest concentration possible in 2016. During the study period, the highest average annual pH values for the station S5 were also in 2016. Turb was the most valuable in 2019. The highest EC value was reported in 2018, although there was relatively little EC value change over the course of the study period. During the years of investigation, the values of the following parameters remained steady or barely changed: TH, Mg, Ca, SO4, TDS, Na, K, and ALK. TSS has the highest possible concentration in 2016, while Cl had the same maximum value in both 2016 and 2018.

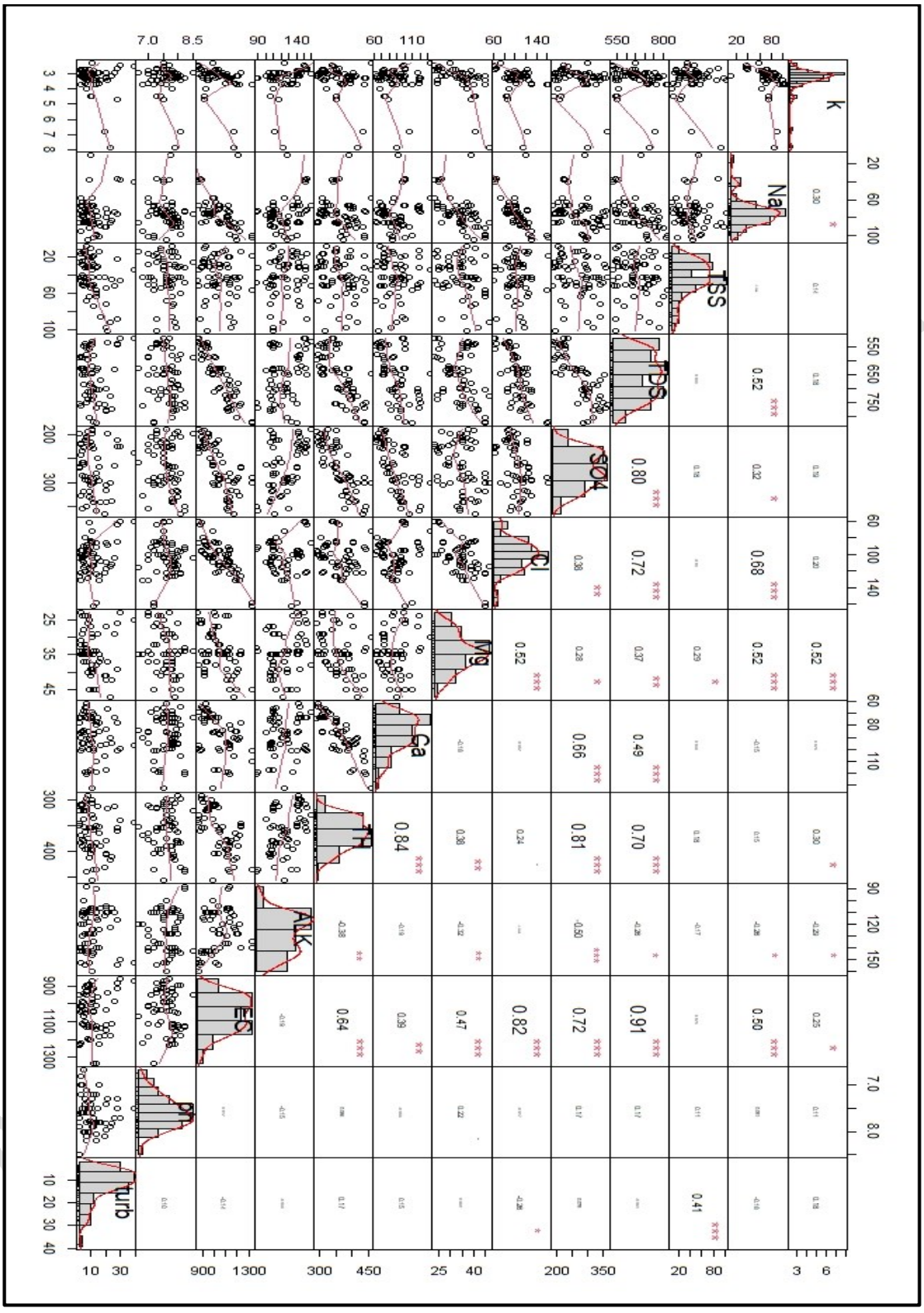


Figure 3: Correlation matrix for the five station of water quality dataset during the study period; The red stars means there is significant relationship.

3.2. The river water quality index (WQI)

The results of average monthly WQI for all sampling stations are shown in Figures 4 to 9. WQI values ranged between (2.38) at the station S3 in April 2018 and (462.49) at the station S1 in August 2019. Similar results were revealed by other researches related at the same study area. Al-Ridah et al. (2020) highlighted extensively the WQI behavior for all sampling stations in Hilla River. WQI values have the lowest value at Al- Hashimyah station in April 2018 and have the highest value at Al-Hesain station in August 2018. The Hilla River water in all stations in 2016 was considered "severely polluted" to "Unfit and unsuitable for humming use". In 2017 and for all study stations there was improvement in WQI in the river water quality compared to 2016, but still WQI is consider "severely polluted". In 2018, the lowest WQI values for all stations was considered the best during the study period, and WQI can be categorized as "Good to Moderately polluted". The WQI value in 2019 was the worst value in all stations during the study period "Unfit and unsuitable for drinking". Also, in 2020 and 2021 the WQI value can be characterized as "severely polluted" (Reza & Singh, 2010) .

The high WQI value of the river was due to the untreated domestic pollution disposal site, which was directly dumped by the lateral outfalls (Reza & Singh, 2010). The water cycle and water quality are significantly impacted by increasing pollution levels, rising water demand, and related increases in pollutant discharges (Whitehead et al., 2006; Whitehead et al., 2009). Because of the combined effect of the decline in rainfall and the rise in potential evaporation as a result of global warming, water of the river has tended to decrease in the recent years. This circumstance indicates that drought might happen more frequently as a result of the effects of global warming (Abdulkareem, 2020).

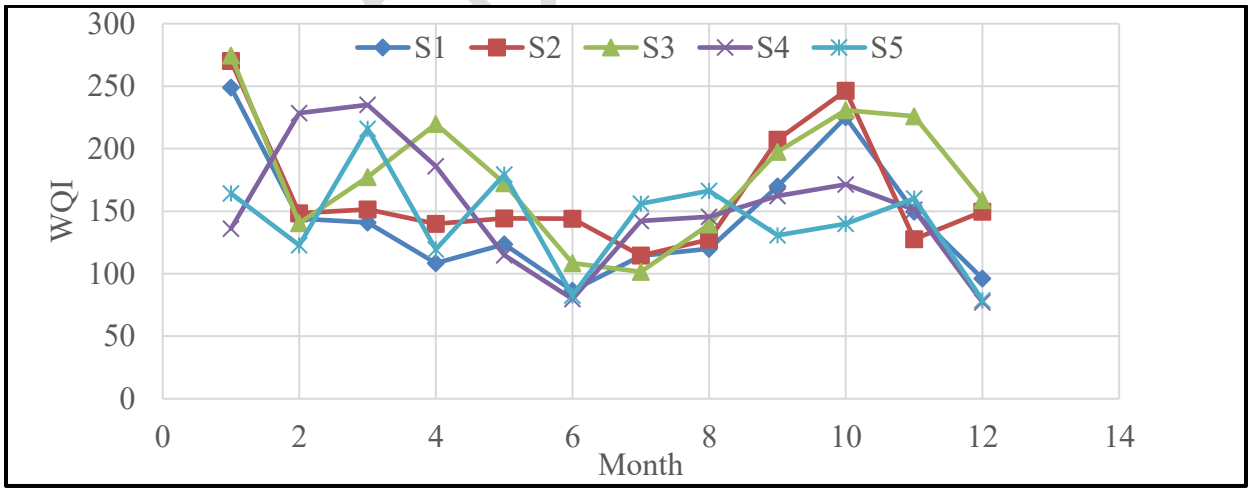


Figure 4. Water quality index for the five stations in 2016.

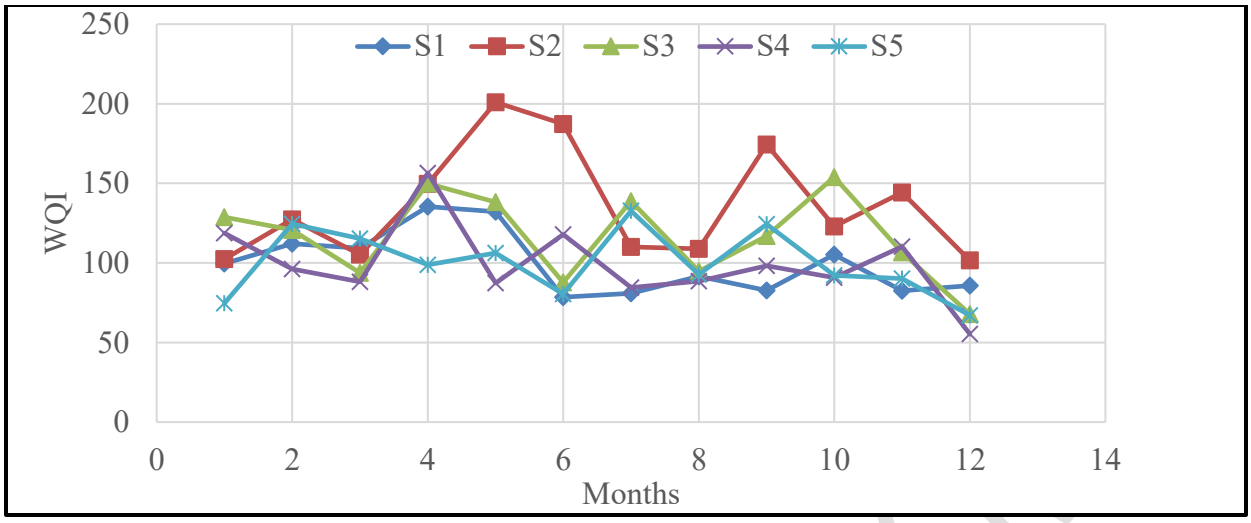


Figure 5. Water quality index for the five stations in 2017.

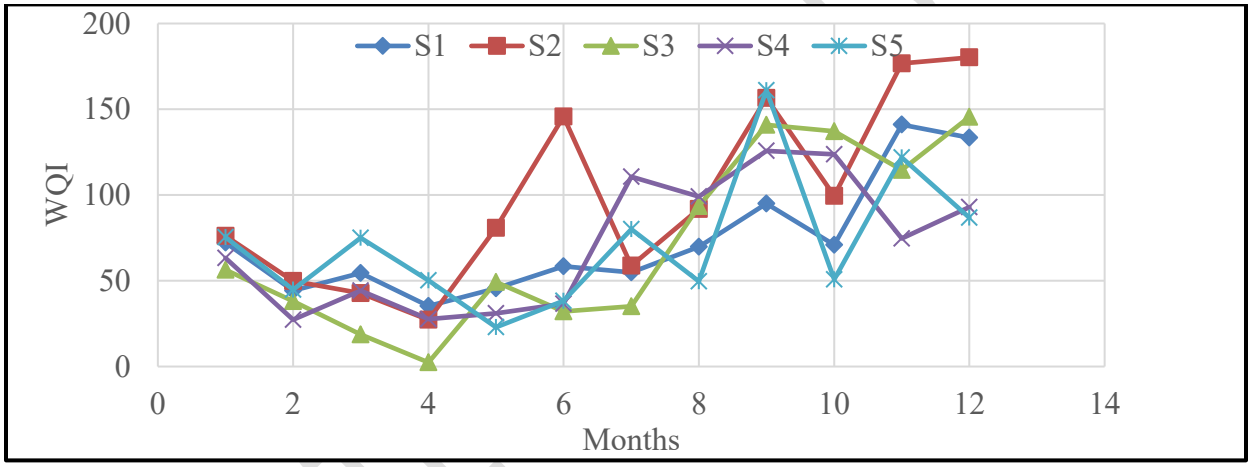


Figure 6. Water quality index for the five stations in 2018.

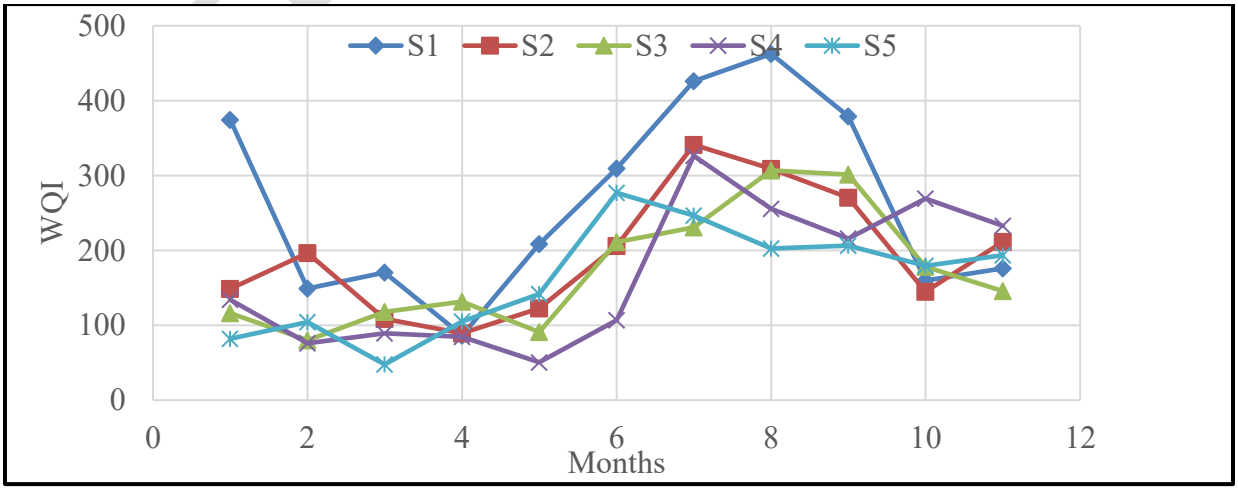


Figure 7. Water quality index for the five stations in 2019.

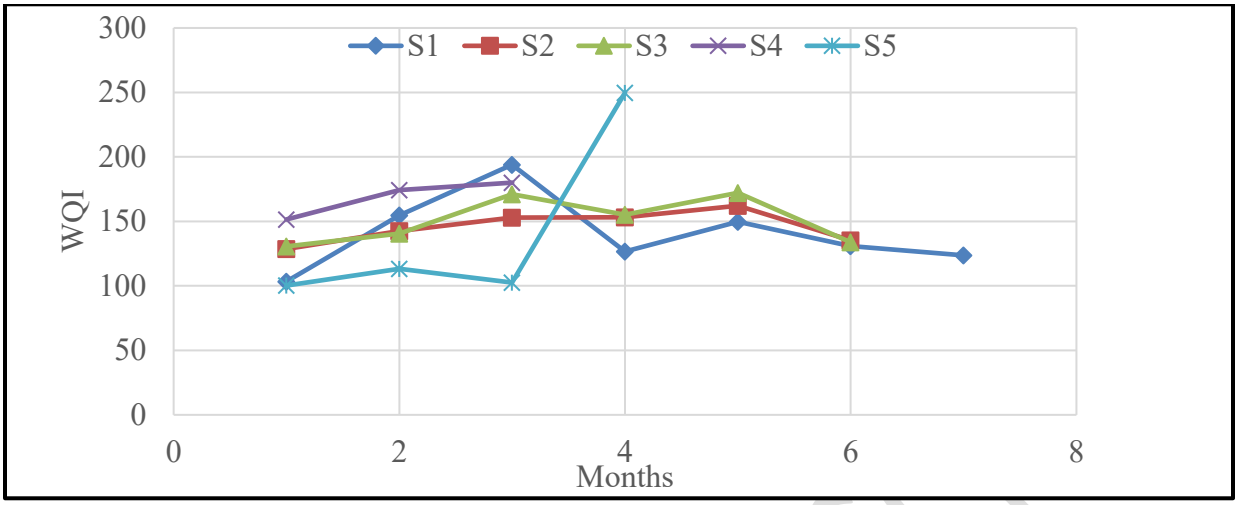


Figure 8. Water quality index for the five stations in 2020.

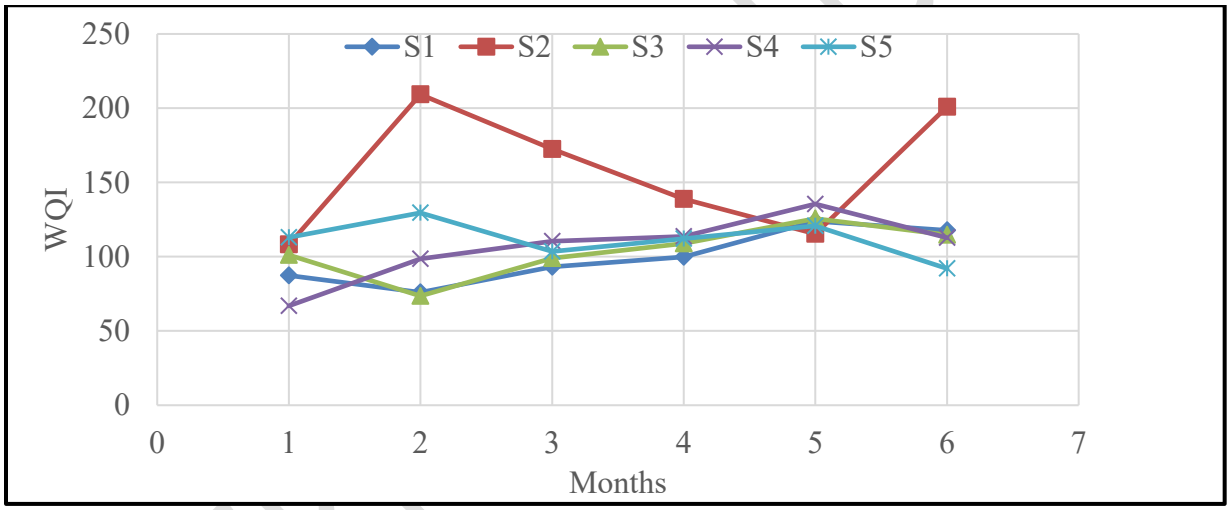


Figure 9. Water quality index for the five stations in 2021.

3.3. The river WQI machine learning analysis

The correlation matrix between the water quality index and other water quality parameters is shown in Figure 10. The thirteen water quality parameters relationships with WQI showed that turbidity has a significant relation with WQI (p-value less than 0.05 and R^2 more than 0.85). As shown in Table 7, testing WQI by Shapiro-Wilk normality test showed that it does not follow the normal distribution. In order to make it normal, WQI values were divided into two groups (less than 220 and more than 220), giving the best valid linear regression model. As a result, the normality test results showed good improvement (Table 8).

For each group, the data was divided into two sets: Train Dataset and Test Dataset. Since there are 291 record of data points from the in-situ measurements used for the WQI calculation. It

was noticed that 264 data points have WQI value of less than 220, and WQI values of more than 220 were 27 data points. Therefore, train data points of WQI<220 are 183 data points, and to validate the result the test data points are 81 points. The correlation matrix for WQI<220 showed there is a significant relationship with turbidity (Figure 11). Also, train data for WQI>220 are 18 date points and test data points were 9 points. The correlation matrix for WQI>220 showed also there is a significant relationship with turbidity (Figure 12).

Table 9 and 10 show the full linear regression model summary statistics for these relationships. Table 9 shows the statistics result of WQI values of less than 220. The linear regression model has a p-value equal to 2.2e-16 and R^2 of more than 70%. Table 10 shows the statistics result of WQI values of more than 220 and the linear regression model that has a p-value of 6.507e-6 and R^2 of more than 86%. Table 11 summarizes the final linear regression models to be used to estimate the Hilla River WQI based on turbidity. The MAE and RMAE were used to validate the difference between the measured values (Test Data) and the model predictions. Results showed that there is good agreement with test data.

Table 7. Shapiro-Wilk normality test for WQI.

W	p-value	Normality case
0.90614	1.724e-12	Not OK

Table 8. Shapiro-Wilk normality test for the WQI groups.

WQI	W	p-value	Normality case
WQI <220	0.99238	0.2159	OK
WQI >220	0.902	0.0529	OK

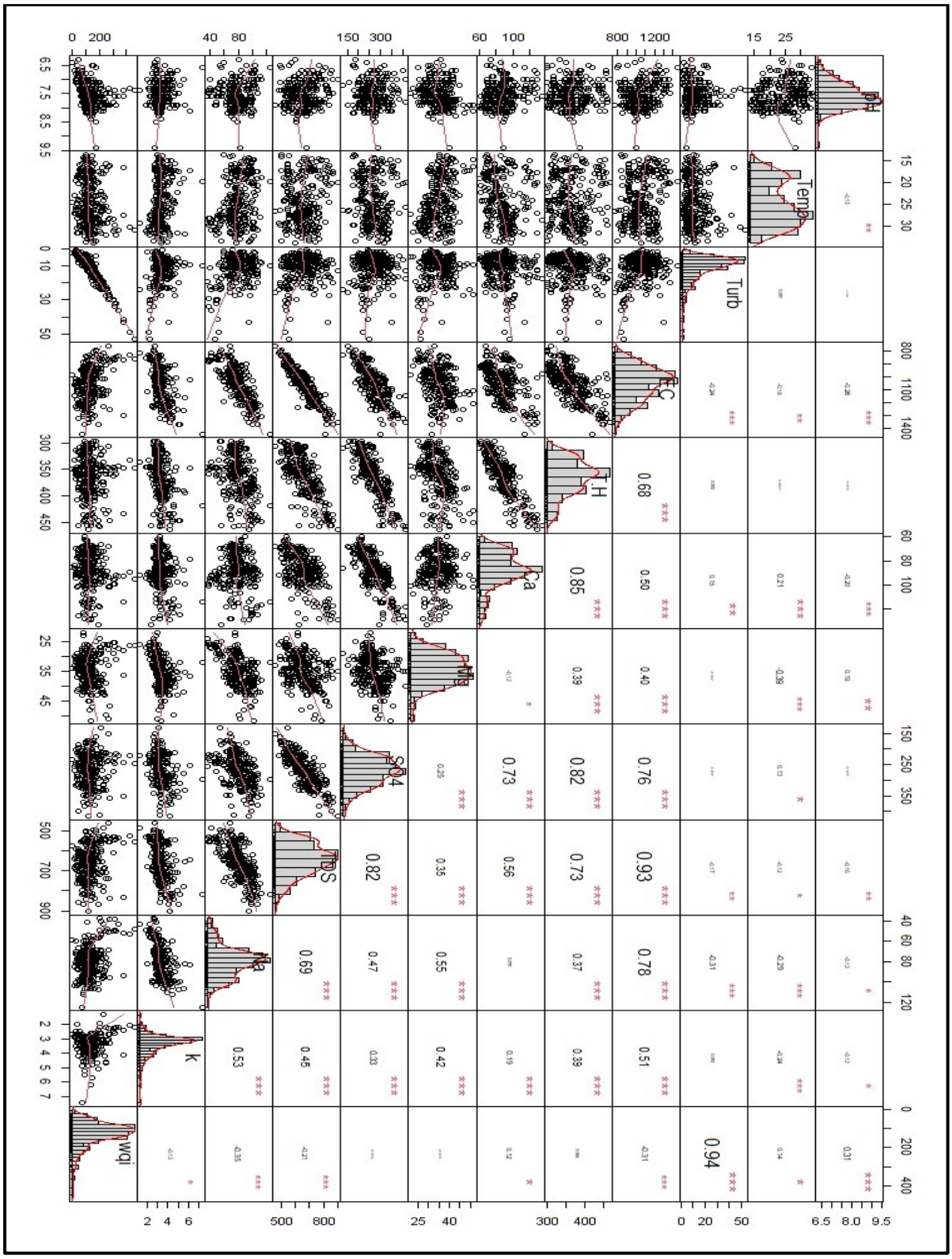


Figure 10. Correlation matrix for the five station of water quality dataset in addition to WQI during the study period; The red stars means there is significant relationship.

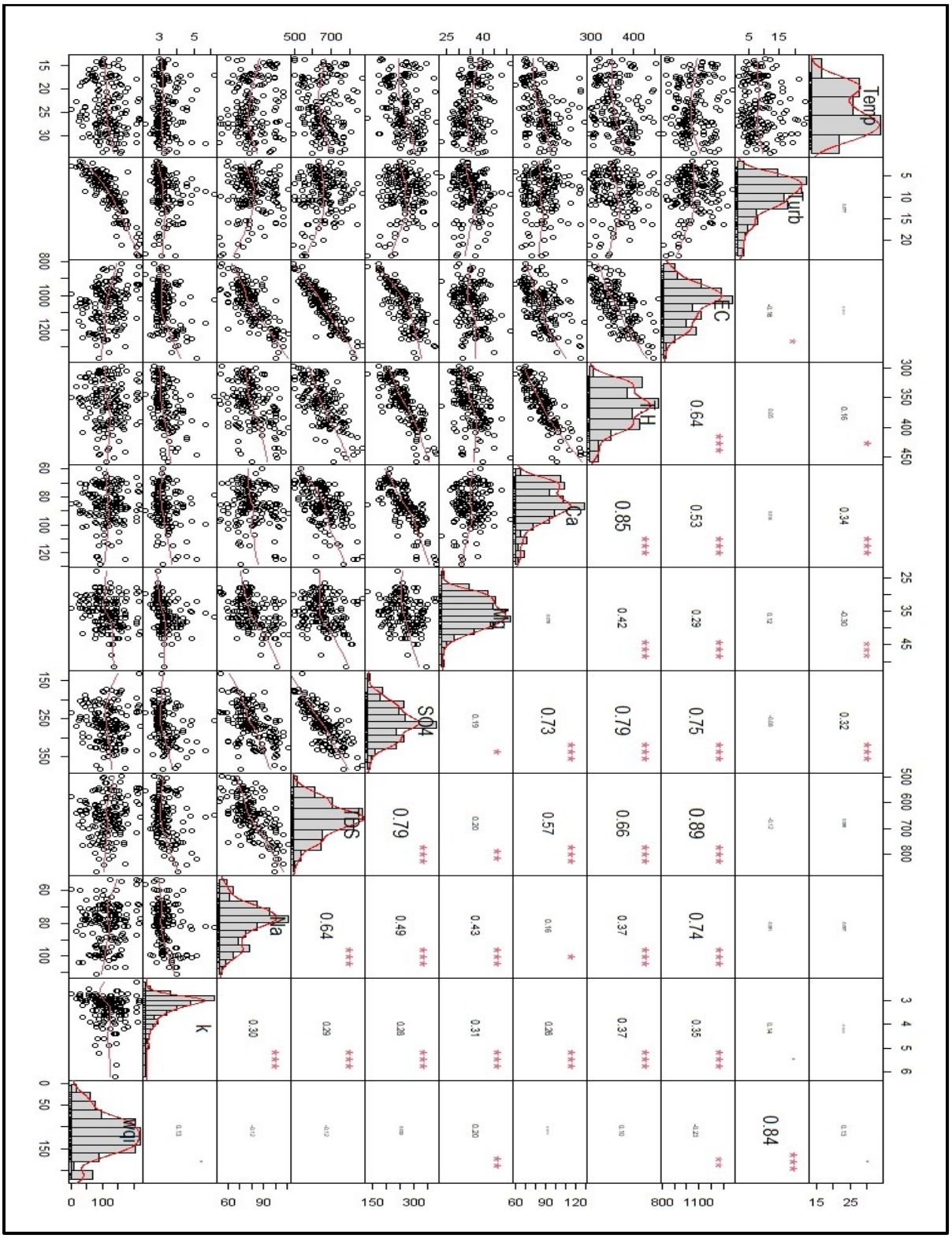


Figure 11. Correlation matrix for the five station of water quality dataset in addition to WQI<220 during the study period; The red stars means there is significant relationship.

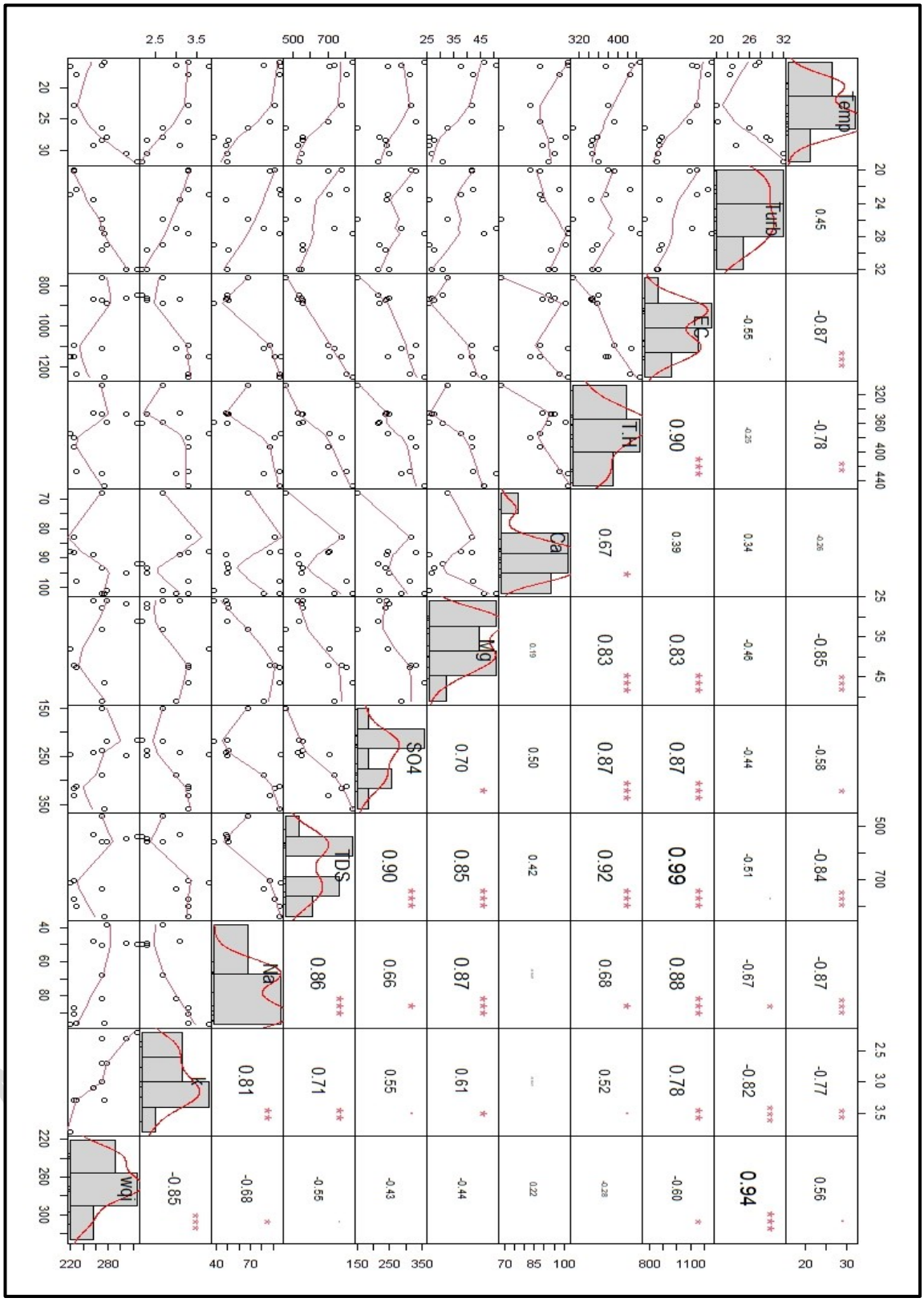


Figure 12. Correlation matrix for the five station of water quality dataset in addition to WQI>220 during the study period; The red stars means there is significant relationship.

Table 9. The full linear regression model summary statistics for the data WQI values of less than 200 with water quality parameters

Residuals:

Min	1Q	Median	3Q	Max
-68.966	-17.195	4.057	13.348	101.682

Coefficients:

	Estimate	Std. Error	t value	Pr (>|t|)
(Intercept)	46.7579	3.8477	12.15	<2e-16 ***
dataf$Turb	7.7177	0.3807	20.27	<2e-16 ***

Residual standard error: 23.89 on 169 degrees of freedom

Multiple R-squared: 0.7086, Adjusted R-squared: 0.7069

F-statistic: 411 on 1 and 169 DF, p-value: < 2.2e-16

Table 10. The full linear regression model summary statistics for the data WQI values of more than 200 with water quality parameters

Residuals:

Min	1Q	Median	3Q	Max
-19.426	-6.982	0.556	7.435	19.059

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	69.3391	22.8720	3.032	0.0126 *
dataf$Turb	7.4363	0.8691	8.556	6.51e-06 ***

Residual standard error: 12.23 on 10 degrees of freedom

Multiple R-squared: 0.8798, Adjusted R-squared: 0.8678

F-statistic: 73.2 on 1 and 10 DF, p-value: 6.507e-06

Table 11. The final linear regression models for WQI.

Model Limitation	Model	MAE	RMSE
WQI<220	WQI=(46.7579)+(7.7177)*(Turb)	17.46917	21.38485
WQI>220	WQI=(69.3391)+(7.4363)*(Turb)	8.989093	12.13835

4. Conclusions

Based on the results obtained, the key conclusions of the present study highlighted the role of machine learning technique in exploring the water quality in surface water. It can be considered a guideline for using an efficient method during the inventory stage for each environmental management process. The main finding of calculating WQI by the weighed arithmetic method for the Hilla River showed that the river can be categorize as "severely polluted". The WQI analysis by the used machine learning algorithm showed there is a significant linear relationship between WQI and turbidity in the river if the water quality parameters were grouped based on WQI of 220. Hence, WQI values of less than 220 are positively correlated with turbidity, and the statistical analysis result showed the developed linear regression model has a p-value of 2.2e-16 and R-squared of more than 70%. In addition, WQI values of more than 220 are positively correlated with turbidity, and the statistical result showed the linear regression model has a p-value of 6.507e-6 and R-squared of more than 86%. Therefore, the river turbidity has a major role in predicting the river health based on WQI.

References

Abbas, N., Wasimi, S., Al-Ansari, N., & Sultana, N. (2018). Water resources problems of Iraq: Climate change adaptation and mitigation. Journal of Environmental Hydrology, 26. http://www.hydroweb.com/protect/pubs/jeh/jeh2018/ansari4.pdf

Al-Bayati, Z. M. K., Nayle, I. H., & Jasim, B. S. (2018). Study of the Relationship Between Spectral Reflectivity and Water Quality Index in Hilla River. International Journal of Engineering & Technology, 7(3.36), 196-200.

Al-Dalimy, S. Z., Al-Zubaidi, H. A. M. (2023). Application of QUAL2K Model for Simulating Water Quality in Hilla River, Iraq. Journal of Ecological Engineering, 24(6), 272-280. https://doi.org/10.12911/22998993/162873

Alobaidy, A. H. M. J., Maulood, B. K., & Kadhem, A. J. (2010). Evaluating raw and treated water quality of Tigris River within Baghdad by index analysis. journal of water resource and protection, 2(7), 629. http://dx.doi.org/10.4236/jwarp.2010.27072

Al-Ridah, Z. A., Al-Zubaidi, H., Naje, A., & Ali, I. (2020). Drinking water quality assessment by using water quality index (WQI) for Hillah River, Iraq. Ecology, Environmental & Conservation, 26, 390-399.

Al-Ridah, Z. A., Naje, A. S., Hassan, D. F., and Al-Zubaid, H. A. M. (2021). Environmental Assessment of Groundwater Quality for Irrigation Purposes: A Case Study of Hillah City in Iraq. Pertanika Journal of Science & Technology. 29(3). https://doi.org/10.47836/pjst.29.3.10

Al-Saadi, B.J.M., Al-Zubaidi, H.A.M. (2024). Numerical simulation of dissolved oxygen and reaeration coefficients in the Hilla River at Saddat Al-Hindiyah Reservoir, Iraq. International Journal of Design & Nature and Ecodynamics, Vol. 19, No. 5, pp. 1801-1808. https://doi.org/10.18280/ijdne.190535

Al-Zubaidi, H. A. M., Naje, A. S., Abed Al-Ridah, Z., Chabuck, A., Ali, I. M., & Shukla, S. K. (2021). A Statistical Technique for Modelling Dissolved Oxygen in Salt Lakes. Cogent Engineering, 8(1). https://doi.org/10.1080/23311916.2021.1875533

APHA. (2017). Standard methods for the examination of water and wastewater. American Public Health Association

Avid Hirst, D., & ob Morris, R. (2001). Water quality of Scottish rivers: spatial and temporal trends. Science of the Total Environment, 265(1-3), 327-342. https://doi.org/10.1016/S0048-9697(00)00674-4

Babu T, Raveena Selvanarayanan, Tamilvizhi Thanarajan & Surendran Rajendran (2024). Integrated Early Flood Prediction using Sentinel-2 Imagery with VANET-MARL-based Deep Neural RNN. Global NEST Journal, 26(10). https://doi.org/10.30955/gnj.06554

Carpenter, S. R., Caraco, N. F., Correll, D. L., Howarth, R. W., Sharpley, A. N., & Smith, V. H. (1998). Nonpoint pollution of surface waters with phosphorus and nitrogen. Ecological applications, 8(3), 559-568. https://doi.org/10.2307/2641247

Chabuk, A., Al-Madhlom, Q., Al-Maliki, A., Al-Ansari, N., Hussain, H. M., & Laue, J. (2020). Water quality assessment along Tigris River (Iraq) using water quality index (WQI) and GIS software. Arabian Journal of Geosciences, 13(14), 1-23. https://doi.org/10.1007/s12517-020-05575-5

Chabuk, A. et al. (2022). Application ArcGIS on Modified-WQI Method to Evaluate Water Quality of the Euphrates River, Iraq, Using Physicochemical Parameters. In: Yang, XS., Sherratt, S., Dey, N., Joshi, A. (eds) Proceedings of Sixth International Congress on Information and Communication Technology. Lecture Notes in Networks and Systems, vol 236. Springer, Singapore. https://doi.org/10.1007/978-981-16-2380-6_58

Chitmanat, C., & Traichaiyaporn, S. (2010). Spatial and temporal variations of physical-chemical water quality and some heavy metals in water, sediments and fish of the Mae Kuang River, Northern Thailand. International Journal of Agriculture and Biology, 12(6), 816-820.

Jegan, D. Surendran, R. & Madhusundar N. (2024). Hydroponic using Deep Water Culture for Lettuce Farming using Random Forest Compared with Decision Tree Algorithm. 8th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2024, pp. 907-914. https://doi.org/10.1109/ICECA63461.2024.10800972

Kankal, N., Indurkar, M., Gudadhe, S., & Wate, S. (2012). Water quality index of surface water bodies of Gujarat, India. Asian J. Exp. Sci, 26(1), 39-48.

Kannan, K. S., & Ramasubramanian, V. (2011). Assessment of fluoride contamination in groundwater using GIS, Dharmapuri district, Tamilnadu, India. International Journal of Engineering Science and Technology, 3(2).

Kotti, M. E., Vlessidis, A. G., Thanasoulias, N. C., & Evmiridis, N. P. (2005). Assessment of river water quality in Northwestern Greece. Water Resources Management, 19(1), 77-94. https://doi.org/10.1007/s11269-005-0294-z

Mohammed, A. A., & Shakir, A. A. (2012). Evaluation the performance of Al-wahdaa project drinking water treatment plant: A case study in Iraq. International Journal of advances in applied sciences, 1(3), 130-138. http://doi.org/10.11591/ijaas.v1.i3.pp130-138

Ochir, A., & Davaa, G. (2011). Application of index analysis to evaluate the water quality of the Tuul River in Mongolia. Journal of Water Resource and Protection, 2011. http://dx.doi.org/10.4236/jwarp.2011.36050

Oko, O. J., Aremu, M. O., Odoh, R., Yebpella, G., & Shenge, G. A. (2014). Assessment of water quality index of borehole and well water in Wukari Town, Taraba State, Nigeria. Assessment, 4(5), 1-9. http://www.iiste.org/Journals/index.php/JEES/article/view/11621/11964

Pathak, S., Prasad, S., & Pathak, T. (2015). Determination of water quality index river Bhagirathi in Uttarkashi, Uttarakhand, India. International Journal of research granthaalayah, 3(9), 1-7. https://doi.org/10.29121/granthaalayah.v3.i9SE.2015.3170

Reza, R., & Singh, G. (2010). Assessment of ground water quality status by using Water Quality Index method in Orissa, India. World Appl Sci J, 9(12), 1392-1397. http://www.idosi.org/wasj/wasj9(12)10/11.pdf

Shanmuga Sundaram, I.K. et al. (2024). Environmental Impact Assessment of Solid Waste Disposal on Groundwater Quality", Global NEST Journal, 26(10). https://doi.org/10.30955/gnj.06080

Sundarapandi, A.M.S. et al. (2024). A Light weighted Dense and Tree structured simple recurrent unit (LDTSRU) for flood prediction using meteorological variables. Global NEST Journal, 26(8). https://doi.org/10.30955/gnj.06242

Venkatraman, M. et al. (2024). Water quality prediction and classification using Attention based Deep Differential RecurFlowNet with Logistic Giant Armadillo Optimization. Global NEST Journal [Preprint]. Available at: https://doi.org/10.30955/gnj.06799

Venkatraman, M, Surendran, R. (2023). Aquaponics and Smart Hydroponics Systems Water Recirculation Using Machine Learning. 4th International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2023, pp. 998-1004. https://doi.org/10.1109/ICOSEC58147.2023.10276310

Whitehead, P. G., Wilby, R. L., Battarbee, R. W., Kernan, M., & Wade, A. J. (2009). A review of the potential impacts of climate change on surface water quality. Hydrological sciences journal, 54(1), 101-123. https://doi.org/10.1623/hysj.54.1.101

Whitehead, P., Wilby, R., Butterfield, D., & Wade, A. (2006). Impacts of climate change on in-stream nitrogen in a lowland chalk stream: an appraisal of adaptation strategies. Science of the total environment, 365(1-3), 260-273. https://doi.org/10.1016/j.scitotenv.2006.02.040