

Water Quality Prediction Modeling of Manchar Lake, Pakistan using Machine Learning Algorithms

Aqsa Zahid*

Department Urban and Infrastructure Engineering, NED University of Engineering and Technology, Karachi, Pakistan

Abstract

Lakes serve as a primary source of fresh water for the local communities and play an important role in improving the environmental well-being of an area. However, around the world, the quality of the lakes is continuously degrading due to various natural and manmade activities. To study the water quality parameters of the lake, this study utilizes Manchar Lake as the study area. The main objective of the research was to investigate the application of machine learning algorithms to predict the water quality index and quality parameters, aiming to overcome the limitations of traditional physical monitoring methods. Multiple machine learning algorithms were evaluated based on their performance measures, including accuracy, precision, recall and F_1 score metrics. The study identified decision tree, random forest and gradient boosting emerging as the most accurate algorithms for predicting the output. These findings highlight the importance of employing advanced machine learning algorithms for timely and accurate assessment of water quality and the development of management and conservation strategies. Such strategies are important to conserve the ecological integrity of freshwater lakes such as Manchar Lake.

Keywords: Main Nara valley drain • Chemical oxygen demand • Particle swarm optimization • Dissolved oxygen

Introduction

Inland water bodies, such as lakes, play a significant role in the enhancement of ecosystem and environmental health, flood control and water storage. They are the main source of fresh water for humans. The condition of these water bodies depends on their catchment size, geographic location, climate and inflow-outflow pattern. However, the quality of these inland lakes is decreasing continuously because of various natural and anthropogenic activities, including siltation, urbanization, agricultural runoff and other human activities. These factors coupled with climate change put immense pressure on the quality and ecological integrity of an area, requiring urgent attention. Therefore, regular monitoring and analyses are important to furnish relevant water quality data to the appropriate authorities for conserving these valuable assets [1-3].

Pakistan is home to numerous freshwater lakes, yet maintaining the quality of its water bodies poses significant challenges, resulting in adverse consequences on the socio-economic dynamics of the country. Among these inland waterbodies, Manchar Lake is the most prominent shallow-water lake in both Pakistan and South Asia. However, since the 1980's, this lake has been designated as an endangered wetland due to continuous degradation in water quality.

Agricultural runoff and domestic discharges that are channeled into the lake through the Main Nara Valley Drain (MNVD) primarily contribute to its contamination. The deteriorating state of the lake highlights the necessity for regular monitoring to conserve it for future utility. Unfortunately, there is a notable absence of a continuous monitoring mechanism in Pakistan to comprehensively assess the lake's condition. Numerous researchers, including, have conducted physical water quality testing and assessments of the Manchar Lake. However, this conventional method is labor-intensive, time-consuming and financially demanding. Moreover, no research has utilized modern AI and machine learning algorithms in assessing water quality in Manchar Lake. Therefore, there is an urgent need to explore innovative tools and data sources that facilitate cost-effective, high-frequency and spatially extensive monitoring of water quality in the lake. This study investigates the various machine learning algorithms to estimate the lake water quality [4,5].

Nowadays, researchers around the world are analyzing water quality parameters by adopting machine learning algorithms. The study by analyzed the five classification methods for Kenya and concluded that the J48 Decision tree had the highest accuracy of 98%. Moreover, advanced artificial intelligence algorithms, including non-linear autoregressive neural networks and Long Short-Term

*Address for Correspondence: Aqsa Zahid, Department Urban and Infrastructure Engineering, sssNED University of Engineering and Technology, Karachi, Pakistan. E-mail: aqsazahid@neduet.edu.pk

Copyright: © 2025 Zahid A. This is an open-access article distributed under the terms of the creative commons attribution license which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

Received: 26 April, 2024, Manuscript No. JCDE-24-133189; **Editor assigned:** 30 April, 2024, PreQC No. JCDE-24-133189 (PQ); **Reviewed:** 14 May, 2024, QC No. JCDE-24-133189; **Revised:** 16 May, 2025, Manuscript No. JCDE-24-133189 (R); **Published:** 23 May, 2025, DOI: 10.37421/2165-784X.2025.15.586

Memory (LSTM) deep learning, were used for analyzing the seven parameters. The algorithms were used to predict the water quality index and concluded that the NARNET model outperformed LSTM in WQI prediction predicted the concentration of Dissolved Oxygen (DO) and Chemical Oxygen Demand (COD) at Liuxi river. The researchers have coupled the Least Squares Support Vector Machines (LSSVM) with Particle Swarm Optimization (PSO) and resulted in high prediction capability. A model for a hypoxic river in south-eastern China was developed. Their study focused on seven parameters and applied various statistical techniques such as multiple linear regression, general regression neural network, Bajick propagation neural network and support vector machine. The research concluded that the model developed using SVM was the most effective and reliable. Further, statistical techniques of Artificial Neural Networks (ANN) and multiple linear regression were used to predict groundwater quality in the Shivganga River Basin of the Western Ghats. The researchers have 34 samples and 13 physiochemical parameters [6-8].

Therefore, the literature review highlights the extensive utilization of machine learning algorithms in predicting water quality parameters worldwide. This shows the applicability, adaptability and effectiveness of these techniques in addressing water quality challenges across diverse geographical locations and environmental conditions.

Study area

Manchar Lake is situated in the district of Jamshoro and Dadu, Sindh province of Pakistan. It lies between 27°40' to 27°50' North latitude and 67°55' to 68°05' East longitude. It is one of the largest freshwater lakes in the country, covering an area of about 250 km² (Figures 1 and 2). However, this total lake area fluctuates seasonally, ranging from 220 km² during dry periods to 256 km² during wet seasons. Moreover, the climate of the area is classified as semi-arid to arid due to hot summers and mild winters. Throughout the year, the area experiences limited precipitation, with an average value of 200 mm/year.

Manchar Lake is surrounded by marshy wetlands, reed beds and fertile agricultural lands which enhances its ecological diversity and richness. Therefore, the lake is an important habitat for various bird species, fish and aquatic plants that provide livelihood opportunities for local communities. Despite its significance as a source of drinking water and irrigation for agriculture, Manchar Lake faces severe water quality challenges due to the diversion of its water to upstream farmlands. The primary inflow of water into the lake originates from catchment areas and the Indus river along with its tributaries. Additionally, the lake receives significant amounts of wastewater effluents from surrounding agricultural lands and villages via the Main Nara Valley Drain (MNVD). This wastewater effluent has adversely impacted its ecosystem, increased salinity levels and posed significant

threats to human health, land productivity, fish populations and the livelihoods of local fishermen and farmers [9].

Data collection

Twelve sampling locations were selected to assess the spatial variation in physiochemical parameters in lake water. Sampling activities were carried out during three seasons in the year 2020: Pre-monsoon (Jan), monsoon (June and July) and post-monsoon (October), for 1.5 years from Jan 2020 to July 2021, to account for potential temporal fluctuations. The assessed physiochemical parameters were: Color, Odor, Temperature, Electrical conductivity (Ec), Total Dissolved Solids (TDS), pH, Turbidity (Turb), Bicarbonate (HCO₃), Chlorine (Cl), Sulfate (SO₄), Calcium (Ca), Magnesium (Mg), Hardness (Hard), Sodium (Na), Potassium(K), Fluorine (F), Nitrate (NO₃) and Alkalis (Alk). Table 1 shows the statistical values (count, mean, standard deviation, min, 25%, 50%, 75% and max) of the input physiochemical variables.

Methods and Materials

Data pre-processing

The Water Quality Index (WQI) is the most effective mathematical tool that consolidates multiple water quality parameters into a single value, providing a comprehensive assessment of overall water quality status. Through various indices, the index reduces the errors encountered from a unilateral perspective and shows the impact of ecosystem variation on water quality. To compute the Water Quality Index (WQI) of Manchar Lake, the collected dataset underwent q-value normalization, transforming it into a scale ranging from 0 to 100. Sixteen parameters were considered and weights were assigned to each parameter for WQI estimation.

The Relative Weight (RW) of physiochemical parameters was determined based on their Assigned Weights (AW), calculated through literature research. Assigned weights ranged from 1.43 to 3.14 for sixteen parameters according to their importance in assessing water quality (Table 2). Parameters with significant health implications were assigned higher weights, as their presence above recommended limits could render the water resource unsuitable for domestic and drinking purposes (Figures 1-3). Equation (1) was used to calculate the relative weighting:

$$\text{Relative Weight} = \frac{\text{Assigned weight of each parameter (AW)}}{\sum_{i=1}^n \text{AW}_i} \quad \text{---(1)}$$

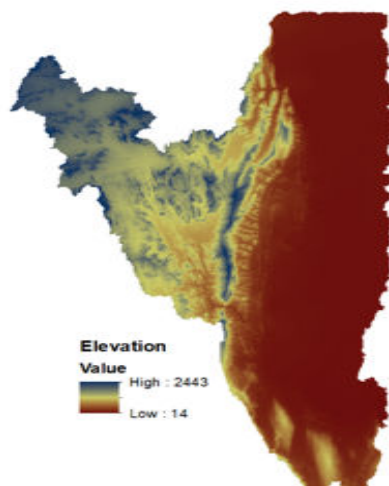


Figure 1. DEM of the Lake.



Figure 2. Manchar lake basin.



Figure 3. Water quality parameters sampling locations.

Where n is the number of parameters. Another parameter required for WQI estimation is the quality rating scale (Q_i) for each parameter. Q_i was calculated by dividing the amount of a particular element in the water sample by the standard concentration of the particulate elements according to the standard set by WHO, as given by the equation (2),

$$\text{Quality rating scale } (Q_i) = C_i / S_i \quad (2)$$

Where, C_i =water quality value of a particular element in the water sample

S_i =water quality value of a particular element obtained from WHO

After the determination of the relative weight and rating scale, the water quality index was calculated. WQI is represented by the multiplication of q -values with their weight (w factor) and then summing them all and dividing by the result of the weighting factors of parameters, as shown in Equation (3) [10-15].

$$WQI = (\sum (q \text{ value} \times W \text{ factor})) / (\sum W \text{ factor}) \quad (3)$$

The World Health Organization drinking water standard was used to calculate the WQI (Tables 1 and 2).

Index	Count	Mean	Std	Min	25%	50%	75%	Max
Temp	19	24.7	3.6	21.3	21.4	21.5	27.6	29.3
Ec	19	7053.2	2027.3	5650	5690	5700	9830	10200
TDS	19	4794.2	1228.8	3699	3724.7	4096	6291	6528
pH	19	7.7	0.2	7.5	7.6	7.7	7.8	8
Turb	19	35	33.5	6.6	7.3	9.4	67.8	88.6
HCO ₃	19	190.5	12.2	180	180	190	200	210
Cl	19	1448.7	477.1	1077	1109	1119	1899	2249
SO ₄	19	1023.2	252	800	804	996	1085	1550
Ca	19	166.5	37.9	136	136	140	220	220
Mg	19	214.5	41.2	171	177.2	201.7	267	279
Hard	19	1266.3	279.9	1070	1080	1080	1650	1700

Na	19	1034.4	319.6	746	749	901	1398.5	1541
K	19	16.6	2.9	14.2	14.8	14.9	20	22
F	19	1.2	0.2	1	1	1	1.4	1.5
NO ₃	19	1.5	0.2	1.2	1.3	1.7	1.7	1.8
Alk	19	3.9	0.3	3.6	3.6	4	4	4.4

Table 1. Statistics of each parameter.

Parameter	Unit	Standard permissible value	Assign weight	Relative weight
Temperature	Degree centigrade	24	3.14	0.096
TDS	mg/l	450	2.93	0.089
pH	-	7.5	2.64	0.08
Turbidity	NTU	5	2.52	0.077
Ec	Micro S/cm	250	3.31	0.101
Nitrate	mg/l	5	2.1	0.064
Total hardness as CaCO ₃	mg/l	100	1.67	0.051
Calcium	mg/l	100	1.71	0.052
Sodium	mg/l	200	1.72	0.052
Magnesium	mg/l	50	1.69	0.052
Flouride	mg/l	1.5	1.61	0.049
Chloride	mg/l	250	1.43	0.044
Total alkalinity	mg/l	120	1.32	0.04
Hydrogen carbonate	mg/l	250	1.47	0.045
Sulphate	mg/l	250	1.72	0.052
Potassium	mg/l	12	1.83	0.056

Table 2. Relative weights of each parameter.

Water quality index classification

Following the computation of the Water Quality Index (WQI), the range and classification of the index were determined based on criteria outlined as depicted in Table 3.

The computed WQI values were then categorized into five classifications, ranging from excellent to unsuitable.

WQI range (%)	Classification
0-25	Excellent
26-28	Good
51-75	Poor
76-100	Very poor
Above 100	Unsuitable

Table 3. Water quality index classification.

Data analysis

Data analysis was comprised of correlation analysis, data splitting and training and machine learning algorithms, adopted for analysis of data [16].

Pearson correlation analysis: Pearson's correlation coefficient was employed to study the relationship between target variables and independent variables. The correlation chart (Table 4) shows the relationship between each independent variable and dependent variable (WQ). The result shows that most of the variables are

negatively correlated, with a strong correlation of pH, turbidity and Total Dissolved Solids (TDS) with water quality.

		Temp	Ec	TDS	pH	Turb	HCO ₃	Cl	SO ₄	Ca	Mg	Hard	Na	K	F	NO ₃	Alk	WQ
1	Temp	1	0.8	0.7	-0.8	0.9	-0.8	0.8	0.6	0.9	0.7	0.8	0.8	0.7	1	-1	-0.6	-0.7
2	Ec	0.8	1	0.9	-0.7	0.9	-0.6	1	0.9	1	1	1	0.8	1	0.8	-0.9	-0.6	-0.8
3	TDS	0.7	0.9	1	-0.8	0.8	-0.6	0.9	0.9	0.9	0.8	0.9	0.6	0.9	0.6	-0.7	0.6	-0.8
4	PH	-0.7	-0.7	-0.7	1	-0.7	0.6	-0.7	-0.5	-0.7	-0.7	-0.7	-0.7	-0.6	-0.7	0.7	0.6	0.7
5	Turb	0.9	0.9	0.8	-0.7	1	-0.7	0.9	0.8	0.9	0.8	0.9	0.9	0.8	0.9	-0.9	-0.5	-0.8
6	HCO ₃	-0.8	-0.6	-0.6	0.6	-0.7	1	-0.6	-0.5	-0.7	-0.6	-0.6	-0.7	-0.5	-0.9	0.8	0.7	0.7
7	Cl	0.8	1	0.9	-0.7	0.9	-0.6	1	0.9	1	0.9	1	0.8	1	0.8	-0.9	-0.2	-0.4
8	SO ₄	0.6	0.9	0.9	-0.5	0.8	-0.5	0.9	1	0.8	0.8	0.9	0.8	0.9	0.7	-0.7	0.8	-0.9
9	Ca	0.9	1	0.9	-0.7	0.9	-0.7	1	0.8	1	0.9	1	0.8	0.9	0.9	-0.9	-0.5	-0.8
10	Mg	0.7	1	0.8	-0.7	0.8	-0.6	0.9	0.8	0.9	1	1	0.8	1	0.7	-0.8	-0.6	-0.9
11	Hard	0.8	1	0.9	-0.7	0.9	-0.6	1	0.9	1	1	1	0.8	1	0.8	-0.9	-0.5	-0.9
12	Na	0.8	0.8	0.6	-0.7	0.9	-0.7	0.8	0.8	0.8	0.8	0.8	1	0.8	0.8	-0.8	-0.3	-0.8
13	K	0.7	1	0.9	-0.6	0.8	-0.6	1	0.9	0.9	1	1	0.8	1	0.7	-0.8	-0.5	-0.7
14	F	1	0.8	0.6	-0.7	0.9	-0.9	0.8	0.7	0.9	0.7	0.8	0.8	0.7	1	-1	-0.6	-0.8
15	NO ₃	-1	-0.9	-0.7	0.7	-0.9	0.8	-0.9	-0.7	-0.9	-0.8	-0.9	-0.8	-0.8	-1	1	0.5	0.7
16	Alk	-0.6	-0.6	0.6	0.6	-0.5	0.7	-0.5	0.6	-0.5	-0.7	-0.6	-0.5	-0.6	-0.7	0.5	1	0.7
17	WQ	-0.9	-0.9	-0.7	0.8	-0.7	0.6	-0.6	-0.8	-0.7	-0.7	-0.8	-0.7	-0.5	-0.7	0.8	0.8	1

Table 4. Parameters correlation chart.

Data splitting: The last step in data pre-processing before the application of machine learning algorithms is data splitting into training and testing datasets. In this study, a split ratio of 70%-30% has been used. For training the model, 70% of the dataset was used while the predictive performance of the model was tested by the remaining 30% dataset.

Machine learning algorithms: The following machine learning algorithms were applied for the water quality prediction analysis:

Random forest: Random forest is a supervised machine learning algorithm that operates on the concept of ensemble learning which consolidates multiple tree classifiers to solve the problem and improve the performance of the algorithm. It develops several tree subsets from the input dataset, gets prediction output from each tree subset and then finally combines these results to get the best output prediction. The greater the number of tree subsets in algorithms, the higher will be its accuracy [17].

Support vector machine: Support Vector Machine (SVM) is a supervised machine algorithm that was proposed by Vapnik. It is as considered the most used algorithm for complex problem-solving related to the classification, learning and prediction of datasets. The algorithm works by representing the dataset parameters as an 'n' number of points and then breaking these points into classes by constructing a hyperplane between them.

Logistic regression: The Logistic Regression Model (LRM) predicts the target variable by developing a relationship between a dependent variable and several independent variables. It then classifies them into several discrete classes. These classes can be used to predict the probability of an observation to occur.

K-nearest neighbor: The K-nearest neighbor algorithm is a supervised machine-learning technique that predicts the classification of unlabeled data by considering both the features and labels of the training data. It classifies datasets by referencing a training model similar to the testing query by utilizing the k nearest training data points (neighbors) that closely resemble the query being assessed. Among the different machine learning algorithms, KNN is one of the simplest techniques for classification due to its adaptive and easily comprehensible design [18].

Decision Tree: The decision tree algorithm is a supervised machine learning method that is mainly used for data mining purposes. Because of its simple structure and accuracy on several data forms, the decision tree method has been used in many implementation fields. It is a hierarchical structure, where each node represents a test feature attribute, each branch shows the outcome of a test and then an outcome prediction is represented by a leaf node. The accuracy of the predicted results depends on the data features and learning decision rules that are used in deriving those results.

XGBoost: The Extreme Gradient Boosting (XGBoost) method was proposed. It is one of the implementations of Gradient Boost Machines (GBM) algorithms for the supervised classification of datasets.

Adaptive Boosting (AdaBoost): Adaptive boosting (AdaBoost) is a machine learning algorithm that works as an ensemble method and can be adapted for predictive modeling techniques. This methodology involves constructing a model that initially assigns equal weights to all data points. Afterward, it adjusts the weights by assigning greater importance to incorrectly classified points. As the process iterates, models are trained repeatedly until a minimized error is achieved, with increased emphasis on points carrying higher weights.

Gradient boosting: Gradient boosting is an ensemble machine-learning technique, that combines several weak train learners into strong train learners. The algorithm updates the weights of the functions by computing the negative gradient of the loss function to the predicted output.

Results and Discussion

This section comprises the performance metrics that were used in the study to assess the performance of the model. All the machine learning algorithms were evaluated based on each of the following performance measures.

Performance metrics are quantitative tools that are used to measure the effectiveness of the machine learning model. The following metrics were used in this study.

Precision: Precision is the ratio between the number of true positives divided by the total number of positive predictions (True positives plus false positives). The higher value leads to a low- false positive rate. The precision was computed by Equation (4).

$$\text{Precision} = \text{True positive} / (\text{True positive} + \text{False positive}) \quad (4)$$

Recall: Recall is the fraction of total positive outcomes to the total positive and false negative outcomes. It was evaluated by equation (5).

$$\text{Precision} = \text{True positive} / (\text{True positive} + \text{False Negative}) \quad (5)$$

Accuracy: Accuracy is the most basic metric, which is the proportion of correctly predicted outcomes to the total observations.

$$\text{Accuracy} = (\text{True positive} + \text{true negative}) / (\text{True positive} + \text{true negative} + \text{false positive} + \text{false negative}) \quad (6)$$

F₁ score: For the classification, the F₁ score is the weighted mean between the precision and recall values. It ranges between 0 to 1, the higher the value, the higher the accuracy. It is given by the Equation (7).

$$F_1 \text{ Score} = 2 \times ((\text{Precision} \times \text{recall}) / (\text{Precision} + \text{recall})) \quad (7)$$

Results of analyzed machine learning algorithms

The analysis of selected machine learning algorithms for the study based on multiple performance metrics shows the distinct patterns for predicting water quality parameters in Manchar Lake. Among the assessed algorithms, the decision tree algorithm resulted in a higher accuracy of 0.9891, coupled with high precision, recall and F₁ score values of 0.9843, 0.9861 and 0.9816, respectively. Similarly, random forest and gradient boosting resulted in better predictive capabilities, with an accuracy of 0.9861 and 0.9741 respectively. In contrast, XGBoost shows comparatively lower performance metrics, with an accuracy of 0.8144 and the lowest F₁ score of 0.7836 among the evaluated algorithms (Table 5). These results highlight the effectiveness of decision tree, random forest and gradient boosting algorithms in predicting water quality parameters for Manchar Lake. The algorithm's results were further analyzed using a graph as shown in Figure 4 [19,20].

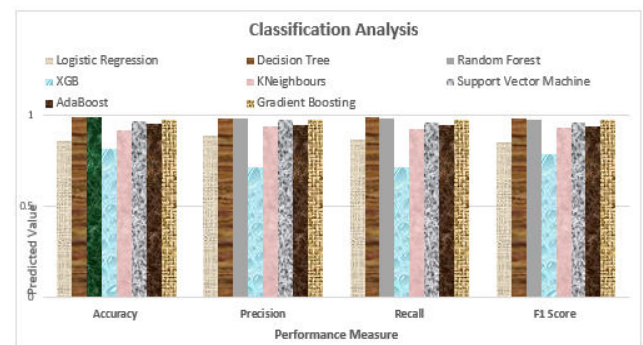


Figure 4. Analysis of performance parameters of each machine learning algorithm.

Algorithm	Accuracy	Precision	Recall	F ₁ score
Logistic regression	0.8571	0.8848	0.865	0.8491
Decision tree	0.9891	0.9843	0.9861	0.9816
Random forest	0.9861	0.9841	0.9789	0.9783
XGBoost	0.8144	0.7161	0.7183	0.7836
KNeighbours	0.9146	0.9383	0.9266	0.9289
Support vector machine	0.9658	0.9736	0.9612	0.9609
Adaboost	0.955	0.9432	0.9473	0.9412
Gradient boosting	0.9741	0.9743	0.9774	0.9739

Table 5. Machine learning algorithms performance measure analysis.

Conclusion

Water is essential for sustaining life on earth, with every human activity directly or indirectly dependent on this vital resource. The socio-economic dynamics of any country depend on how well the country manages its available resources. Pakistan, being home to several freshwater lakes, that support human health, agriculture, industry and ecosystem functioning, faces several challenges in maintaining its water quality. The continuous threat from urbanization, climate change and population expansion contributes to the degradation of the lake. Therefore, the analysis and prediction of water quality before its utilization has become a prerequisite in the present time. For Manchar Lake, maintaining water quality is of paramount importance due to its significance as a crucial water source for local communities and supporting various ecological and economic activities.

The effectiveness of machine learning algorithms in predicting water quality parameters for Manchar Lake highlights their importance for conserving the resource. By adopting advanced predictive modeling techniques, respective departments and stakeholders can proactively address water quality challenges and the sustainability of Manchar Lake for current and future generations.

References

- Xiang, Yunrong and Liangzhong Jiang. "Water Quality Prediction Using LS-SVM and Particle Swarm Optimization." *IEEE* (2009): 900-904.
- Aldhyani, Theyazn H. H., Mohammed Al-Yaari, Hasan Alkahtani and Mashael Maashi. "Water Quality Prediction Using Artificial Intelligence Algorithms." *Appl Bionics Biomech* (2020): 6659314.
- Gakii, Consolata and Jennifer Jepkoech. "A Classification Model for Water Quality Analysis Using Decision Tree." (2019).
- Tang, KL, FL Zheng, KL Zhang and BK Wang, et al. "Research Subjects and Methods of Relationship between Soil Erosion and Eco-Environment in the Ziwuling Forest Area." 17 (1993): 3-10.
- Afifah Tarmizi, Afifah Tarmizi, AN Ahmed and A El-Shafie. "Dissolved oxygen prediction using Support Vector Machine in Terengganu River." (2014): 2182-2188.
- Walkley, Aldous and I. Armstrong Black. "An Examination of the Degtjareff Method for Determining Soil Organic Matter and a Proposed Modification of the Chromic Acid Titration Method." *Soil Sci* 37 (1934): 29-38.
- Lerios, Jefferson L and Mia V. Villarica. "Pattern Extraction of Water Quality Prediction Using Machine Learning Algorithms of Water Reservoir." *Int J Mech Eng Robot Res* 8 (2019): 992-997.
- Ji, Xiaoliang, Xu Shang, Randy A. Dahlgren and Minghua Zhang. "Prediction of Dissolved Oxygen Concentration in Hypoxic River Systems Using Support Vector Machine: A Case Study of Wen-Rui Tang River, China." *Environ Sci Pollut Res Int* 24 (2017): 16062-16076.
- Kadam AK, VM Wagh, AA Muley and BN Umrikar, et al. "Prediction of Water Quality Index Using Artificial Neural Network and Multiple Linear Regression Modelling Approach in Shivganga River basin, India." *Model Earth Syst Environ* 5 (2019): 951-962.
- Khan, Yafra and Chai Soo See. "Predicting and Analyzing Water Quality Using Machine Learning: A Comprehensive Model." *IEEE* (2016): 1-6.
- Imran, Uzma, Asmat Ullah, Rasool Bux Mahar and Kaleemullah Shaikh. "Distribution, Source Identification and Ecological Risk Assessment of Selected Trace Elements in Sediments from Manchar Lake, Pakistan." *Soil and Sedi Inter J* 32 (2023): 615-635.
- Imran, Uzma, Asmat Ullah, Rasool Bux Mahar and Kaleemullah Shaikh, et al. "An Integrated Approach for Evaluating Freshwater Ecosystems Under the Influence of High Salinity: A Case Study of Manchar Lake in Pakistan." *Environ Monit Assess* 195 (2023): 1340.
- Toth, SJ and AL Prince. "Estimation of Cation-Exchange Capacity and Exchangeable Ca, K and Na Contents of Soils by Flame Photometer Techniques." *Soil Sci* 67 (1949): 439-446.
- Kumar, L, R Kumari, A Kumar and IA Tunio, et al. "Water Quality Assessment and Monitoring in Pakistan: A Comprehensive Review." *Sustainability* (2023).
- Zhao C, N Shao, S Yang and H Ren, et al. "Integrated Assessment of Ecosystem Health Using Multiple Indicator Species." *Ecol Eng* 130 (2019): 157-168.
- Das Kangabam, Rajiv, Sarojini Devi Bhoominathan, Suganthi Kanagaraj and Munisamy Govindaraju. "Development of a Water Quality Index (WQI) for the Loktak Lake in India." *Appl Water Sci* 7 (2017): 2907-2918.
- Olsen, Sterling Robertson. "Estimation of Available Phosphorus in Soils by Extraction with Sodium Bicarbonate." (1954).
- Srivastava, Garima and Pradeep Kumar. "Water Quality Index with Missing Parameters." *Int J Eng Res* 2 (2013): 609-614.
- Yadav, Ashok Kumar, Parveen Khan and Sanjay K Sharma. "Water Quality Index Assessment of Groundwater in Todaraisingh Tehsil of Rajasthan State, India-A Greener Approach." *J Chem* 7 (2010): S428-S432.
- Yidana, Sandow Mark and Adadow Yidana. "Assessing Water Quality Using Water Quality Index and Multivariate Analysis." *Environ Earth Sci* 59 (2010): 1461-1473.

How to cite this article: Zahid, Aqsa. "Water Quality Prediction Modeling of Manchar Lake, Pakistan using Machine Learning Algorithms." *J Civil Environ Eng* 15 (2025): 586.