Development Of Water Quality Prediction Model For Narmada River Using Artificial Neural Networks

Shubham Lakhera (shubhamlakhera9179@gmail.com)
JEC Jabalpur: Jabalpur Engineering College https://orcid.org/0000-0001-8204-256X
Sunayana Chandra
CSIR-NEERI: National Environmental Engineering Research Institute CSIR
Dal Chand Rahi
JEC Jabalpur: Jabalpur Engineering College

Research Article

Keywords: Artificial neural network, Water quality index, Water quality prediction, Artificial intelligence, Machine learning

Posted Date: December 29th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1166542/v1

Title: - Development Of Water Quality Prediction Model For Narmada River Using Artificial Neural Networks

Author: - Shubham Lakhera

Affiliation: - Jabalpur Engineering College, Department of Civil Engineering, Jabalpur, Madhya Pradesh, India.

Email: - slakhera.ce19me14@jecjabalpur.ac.in

ORCID-ID: - 0000-0001-8204-256X

Co-Author: - Dr. Sunayana

Affiliation: - Scientist, National Institute of Environmental Engineering Research, Delhi Zonal Centre (CSIR-NEERI)

Email: - nayanahbti@gmail.com

Co-Author: - Dal Chand Rahi

Affiliation: - Jabalpur Engineering College, Department of Civil Engineering, Jabalpur, Madhya Pradesh, India.

Email: - dcrahi76@gmail.com

Address for Correspondence: - Lakhpati Traders, Near Amar Engineering Works, Jail Road, Khurai, 470117, Dist. Sagar, Madhya Pradesh, India.

Conflict of Interest Statement :

The authors whose names are listed immediately below certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

Author names:

Shubham Lakhera

Dr. Sunayana

D.C. Rahi

DEVELOPMENT OF WATER QUALITY PREDICTION MODEL FOR NARMADA RIVER USING ARTIFICIAL NEURAL NETWORKS

Abstract:

The lack of a universal system for analysis, prediction, and storage of water quality and condition of rivers in Madhya Pradesh has led to uneven policy-making and poor management ultimately posing issues in health, irrigation and keep increasing pollution in rivers. This study is a part of developing a central system for river water quality assessment and prediction. The conventional method of water quality assessment is based on the calculation of the water quality index which can be very complex and time-consuming. This paper aims to develop a water quality prediction model with the help of an Artificial Neural Network (ANN) for predicting the water quality of the Narmada River using two machine learning algorithms Levenberg and Gradient Descent and the results were compared. This research uses the surface water historical data of years 2018, 2019 of the river Narmada with monthly time intervals. Data is obtained from the Central Pollution Control Board resource called Narmada Automatic Sampling Collection Stations System. For training the network 10 water quality parameters including, DO, BOD, Turbidity, pH, etc. After training the networks were accessed using Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Coefficient of Correlation (R) out of which 2 best performing networks with 7 (Training R = 0.80083, Testing R = 0.5767) and 19 (Training R = 0.6594, Testing R = 0.7424) Neurons in the hidden layer, were selected from Levenberg algorithm and, 5 (Training R = 0.7670, Testing R = 0.8123) & 17 (Training R = 0.8631, Testing R = 0.8981) Neurons in the hidden layer were selected from Gradient descent algorithm. This simplifies the calculation of WQI take care if any sampling station is out of service and data is not available for some reason. Further, the aim is to refine the prediction location-wise to be able to make a better decision when & where to implement the measures to reduce the pollution or the knowledge level of treatment required to make the water fit for use beforehand. This would be helpful in the treatment of water for use in Domestic or Irrigation Purposes.

***Keywords* - Artificial neural network, Water quality index, Water quality prediction, Artificial intelligence, Machine learning.**

I. INTRODUCTION

Humans have polluted water resources both surface and subsurface (Ahuja 2009) so much that almost everywhere not only for drinking but also for use in moderate industry, purification of water has become necessary. Earth has a lot of water but not all of it is readily available which can be used directly, about 0.06% is easily accessible (Ahuja 2013). This problem is getting worse in the developing countries, 37.7 million Indians are affected by drinking polluted water, over 1.5 million children die because of diarrhoea every year. The use of Machine Learning and Artificial Intelligence is increasing day by day in every field, with the tools available to us we can make this water crisis less problematic. In this paper, using Artificial Neural Network a model was developed for better understanding and prediction of River Narmada's water. It used 9 parameters and predicted the water quality index. The conventional method of water quality assessment is based on the calculation of the water quality index which can be very complex and time-consuming. This simplified the calculation of WQI take care if any sampling station is out of service and data is not available for some reason. Further, the aim is to refine the prediction location-wise to be able to make a better decision when & where to implement the measures to reduce the pollution or the knowledge level of treatment required to make the water fit for use beforehand using Artificial Neural Networks. This would be helpful in the treatment of water for use in Domestic or Irrigation Purposes.

II.RESEARCH METHODOLOGY

2.1 NARMADA RIVER

Narmada river is one of the most important rivers in central India. It is not only the backbone of many local businesses but also has great cultural importance. The basin extends over 98,976 km^2 in Madhya Pradesh, Gujarat, Maharashtra, and Chhattisgarh but mostly in plateau tracts of the peninsular region of India. River Narmada originates as a sub-surface spring at Amarkantak on the Madhya Pradesh-Chhattisgarh border. The River flows westwards for about 1300 km to join the Arabian Sea near Bharuch in Gujarat (Narmada Valley Development Authority, 2013). The River has a utilizable surface water resource of about 34,500 million cubic meters. The Narmada basin coordinates are 72° 38' to 81° 43'E (Longitude) 21° 27' to 23° 37'N (Latitude), hemmed between Vindya and Satpura ranges. It is the longest West flowing River in India. The origin of the River and the boundaries of the basin are of special importance as Amarkantak marks the boundary between the Narmada and the Ganga basins. Narmada Jayanti which is a festival celebrated by the people of Jabalpur worshipping river Narmada to bring peace and prosperity in their life. (Clean Ganga Report, Gov of India P 120-121).

2.2 DATA COLLECTION

The data is obtained from the website of the Central Pollution Control Board website. Over the course of years, CBPC has installed "Automatic River Sampling Stations" on the river which measure the quality of the river and send the data back to the server where it is processed and both raw data and the state of the river is made available. One of the drawbacks of this system is that it does not take account of any out-of-service stations or missing values. The river water quality information is very poor providing only if the river water is 'satisfactory' or 'unsatisfactory' or use. Over 1200 values were used to train the model. The dataset contained the following 27 parameters: Temperature, Turbidity, Colour, Odour, pH, Spatial Conductivity, Total Solids, Dissolved Solids, Suspended Solids, Ammonia Nitrogen, Nitrate Nitrogen, Nitrate Nitrogen, Phosphate (PO_4), Chloride, Sulphate (SO_4), Total Alkalinity, Total Hardness, Calcium Hardness, Magnesium Hardness, Dissolved Oxygen, Biochemical Oxygen Demand$_3$, Chemical

Oxygen Demand, Sodium, Potassium, Total Coliform, Faecal Coliform, and Total Kjeldahl Nitrogen. For our study, 10 parameters were selected for training the model see Table 9. The Summary of Dataset is given in Table 1.

2.3 RELATED STUDIES

Several studies, on analysis and monitoring of water quality, have been done. Methodologies range from statistical techniques, visual modeling, prediction algorithms, and decision making. Multivariate statistical techniques like Principal component Analysis (PCA) have been used to determine the relationship among different water quality parameters Tripathi & Singal (2019). Wechmongkhonkon (2012), utilizes a multilayer perceptron neural network through Levenberg-Marquardt algorithms to group the water nature of Dusit District canals of Bangkok, Thailand. The result demonstrates that the neural network achieves well with a high accuracy order rate of 96.52%. Xiang and Jiang (2009), found that through simulation testing the Least square support vector machine with particle swarm optimization method show high proficiency in estimating the water quality of the Liuxi River. Khan & Soo see (2016), devised a comprehensive methodology using Artificial Neural Networks with Nonlinear Autoregressive (NAR) time series model that analyses and predicts water quality of Island park village, situated in the South-Western Nassau County New York. In their study four parameters i.e. Chlorophyll, Specific Conductance, Dissolved Oxygen and Turbidity were used. In the field of water and wastewater technology, there have been many studies and predictions from these models that have become better and better.

The paper published by Thikra, (2021) on prediction of level on contamination in a water distribution system showed how neural networks can handle problems complex problems efficiently. Dawood & Nayak (2021) comparing, results from the Levenberg Marquardt algorithm and Scaled Conjugate Gradient algorithm were compared of Godavari river. Najah (2013) prediction of total dissolved solids, electrical conductivity, and turbidity. Nouraki (2021) Predicting the level of total dissolved solids, sodium absorption ratio, and total hardness using various machine learning methods such as multiple linear regression, M5P model tree, support vector regression, random forest regression and comparing their results . Salari (2021) research on Application of SVM and FFBP for prediction of water quality in wetlands. Wagh (2016) predicting groundwater suitability for irrigation, values of sodium adsorption ratio (SAR), residual sodium carbonate, magnesium adsorption ratio, provide a good understanding on how to follow up with this field of implication of Machine Learning in Environmental Engineering. Vijay (2021) study pointed out the performance of various functions like Tanh, Maxout, and rectifier in groundwater of vellore district. Hmoud (2021) developed an Adaptive neuro-fuzzy interference system (ANFIS) for the prediction of water quality index using feed-forward neural networks (FFNN) and K-nearest neighbors as classifiers. Their model showed a very high R-value of 92.39 in the testing phase.

2.4 WATER QUALITY INDEX

Water Quality Index is a tool that helps up in the management of water quality by easily evaluating and processing large water quality datasets. WQI models are based on aggregation functions which allow an analysis of large temporally and spatially-varying water quality datasets to produce a single value, that is water quality index (Uddin 2021).

The process of calculation of WQI comprises four steps.

1. The Water Quality parameters of our significance are selected.
2. Each water quality parameter is then converted to a single-value dimensionless sub-index.
3. The weighting factor for each water quality parameter is determined.
4. Finally using an aggregation function on sub-indices, WQI is calculated.

In supervised machine learning, the labeled dataset is required. There are many methods available by which WQI can be calculated, one of the first developed by Horton (Horton et al. 1965) and Brown (1970), since then many changes have been made and many new models have been developed. After a comprehensive comparison of these models, keeping in mind the parameters required to calculate the index, and parameters available in our dataset, the weighted arithmetic water quality index method was selected to compute the WQI value (Table 8 & 9).

2.4.1 Weighted Arithmetic Water Quality Index Method

There are many methods available to calculate the water quality index. Brown index, National Sanitation foundation index, Smiths Index, Horton Index. For this study, the weighted arithmetic water quality index (Aldhyani et al. 2020) method is used for calculating the WQI value (See Table 9). In this, all 10 parameters were included. WQI is calculated by the following steps:

Step 1: Collect data of various physio-chemical water quality parameters.

Step 2: Calculate Proportionality constant 'k' value using formula:

$$k = \left(\frac{1}{\frac{1}{\Sigma_{i=1}^{n} S_i}}\right) \tag{1}$$

where 'S_i' is standard permissible for the n^th^ parameter. (See Table 3.)

Step 3: Calculate quality rating for the n^th^ parameter (q_n) where there are n parameters. This is calculated using the formula.

$$q_n = 100 \left(\frac{(V_n - V_{ideal})}{(S_n - V_{ideal})}\right) \tag{2}$$

where S_n = Estimated value of the nth parameter of the given sampling station. V_i = Ideal value of the n^{th} parameter in pure water. And S_n = Standard permissible value of the n^{th} parameter.

Step 4: Calculate unit weight for the nth parameters.

$$w_n = \left(\frac{k}{S_n}\right) \tag{3}$$

Step 5: Calculate Water Quality Index (WQI) using formula. (See Table 8)

$$WQI = \left(\frac{\sum W_n q_n}{\sum W_n}\right) \tag{4}$$

2.5 ARTIFICIAL NEURAL NETWORK

An artificial neural network is a digital copy of a biological neural network. It consists of input, output, and hidden layers. These layers consist of neurons or nodes which are interconnected with each other by weights combining many forms a network (see

Figure 12). A training algorithm is used for training and optimization which is accomplished by minimizing the error or loss function & using a transfer function, the predictions are transferred to the output.

Multilayer perceptron introduced by Rosenblatt 1958 (network containing many nodes/neurons) is feed-forward neural networks of multiple layers trained by any standard backpropagation algorithm. The objective is to learn how to transform input data into the desired response. The perceptron computes a single output from multiple real-valued inputs by forming a linear combination according to its weights and then putting the output through some nonlinear activation function. To build any multilayer perceptron, the number of hidden layers and neurons in the network needs to be calculated. The number of neurons and hidden layers depends upon many factors like the amount of noise in the dataset, complexity of the function, training cases. Using too few neurons would results in poor performance of the network, using too much would just make the network memorize the values and not learn anything. Following are the rules given based on previous studies:

- The number of hidden neurons should be between the size of the input layer and the size of the output layer (Blum 1992).
- The number of hidden neurons should be 2/3 of the (size of the input layer + size of output layer).
- The number of hidden neurons should be less than twice the size of the input layer (Berry and Linoff, 1977).
- The number of hidden neurons should be equal to dimensions (principal components) needed to capture 70-90% or the variance (spread of data from the mean) of the input dataset (Boger & Guterman, 1997).

The selection training & selection process is summarized in Figure 1. To find the best performing network, over 50 networks were trained first by Levenberg and then by Gradient descent, each was accessed for MSE, RMSE & R. Out of these, 2 networks one which showed the lowest MSE in the training phase, second which showed the Highest R were selected, and their performance is shown in Figure 2-11.

2.5.1 Evaluating Model Accuracy

Evaluating the accuracy and performance of the model is an integral part of machine learning. Mean absolute error, mean squared error, Root mean squared error, and R^2 are mainly used to evaluate the prediction error rates and performance in regression analysis (Chicco et al. 2021). Recent findings suggested RMSE varies with the variability of the error magnitudes and sample size *n*. In this study, MSE, R is preferred for the selection of best performing model (Willmott & Matsurra 2005).

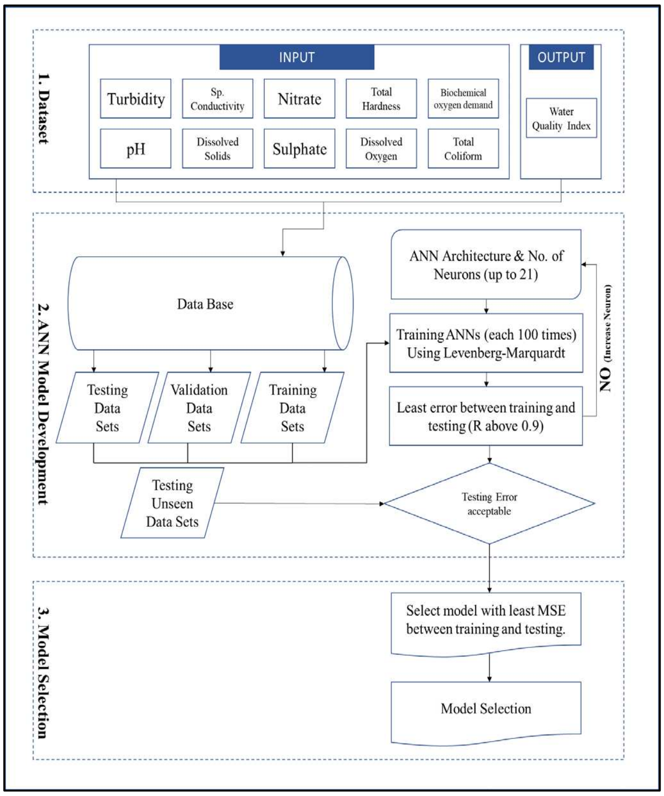


FIGURE 1: ARCHITECTURE OF THE NETWORK

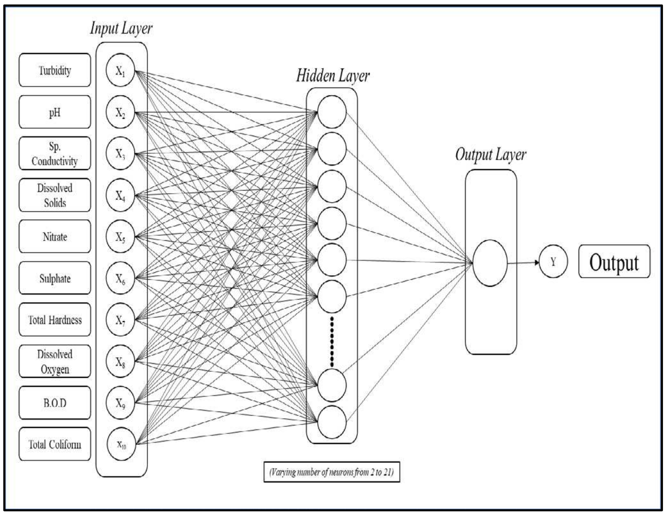


FIGURE 12: NETWORKS TRAINING AND SELECTION PROCESS

Here: X_i – Predicted value, Y_i – Actual Value

- MSE – Mean squared error represents the difference between the original and predicted values extracted by squared average difference over the data set. Very good for detecting outliers.

$$MSE = \frac{1}{m}\sum_{i=1}^{m}(X_i - Y_i)^2 \tag{5}$$

- RMSE – Root mean squared error is the error rate by the square root of MSE. (Best value = 0)

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(X_i - Y_i)^2} \tag{6}$$

- R^2 – R^2 also called the *coefficient of determination* (Niles 1992), is the measure of variance in the dependent variable predictable from the independent variables.The value ranges from -ꝏ to +1 (best value = +1). R^2 is the measure of how well a model can predict the data, the higher the value of R^2, the better the model is predicting the data.

$$R^2 = 1 - \frac{\sum_{i=1}^{m}(X_i - Y_i)^2}{\sum_{i=1}^{m}(\bar{Y} - Y_i)^2} \tag{7}$$

- R – Also called *'correlation coefficient'* is the measure of linear dependence between 2 variables. There are many methods of calculating '*R*', Karl Pearson's coefficient of correlation, Spearman's rank correlation coefficient, Method of least squares. Using MATLAB's *plotregression, R* was calculated in training, testing, validation, and on unseen data.

$$\rho(A, B) = \frac{1}{N - 1} \sum_{i=1}^{N} \left(\frac{A_i - \mu_A}{\sigma_A}\right) \left(\frac{B_i - \mu_B}{\sigma_B}\right) \tag{8}$$

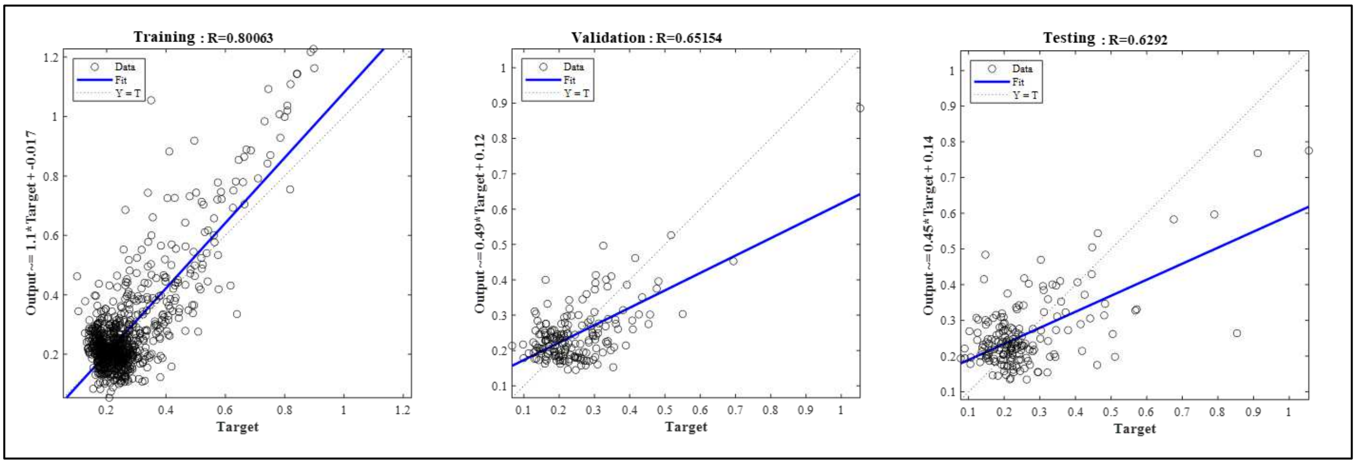
Where, σ_A & σ_B standard deviation and μ_A & μ_B are the mean of A & B respectively.

III. RESULTS & DISCUSSION

3.1 Results

Levenberg algorithm performed well, 2 best models with 7 and 19 Neurons were finally selected which had R values of 0.80063, 0.6594 in training & 0.6292, 0.7424 in testing respectively. Figure 13 & 14 shows the, Predicted and Actual values from Levenberg algorithm. The dataset of unseen inputs was also run through the trained models, from which 5 and 19 Neuron models showed the best results with R values of 0.7266 & 0.7815 respectively (See Table 10).

FIGURE 2: COEFFICIENT OF REGRESSION (R) PLOT OF 7 MODEL TRAINING VALIDATION TESTING.



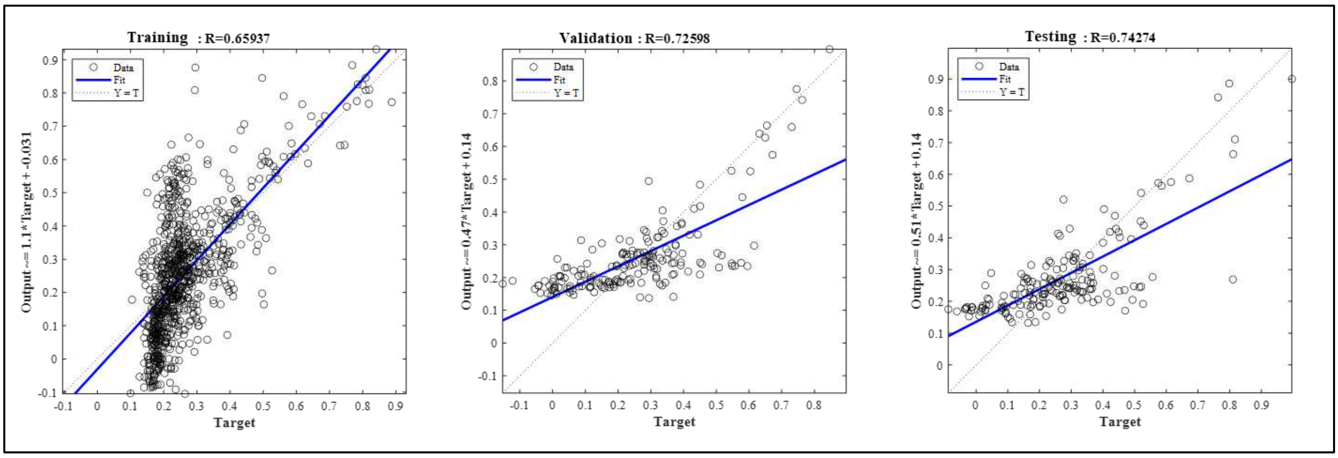


FIGURE 4: PERFORMANCE OF 7 NEURONS MODEL

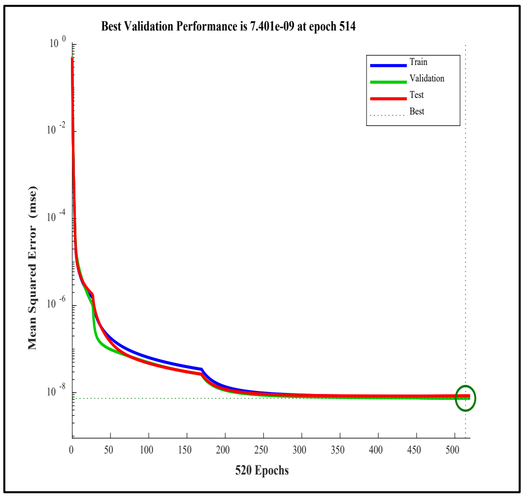


FIGURE 5: PERFORMANCE OF 19 NEURONS MODEL

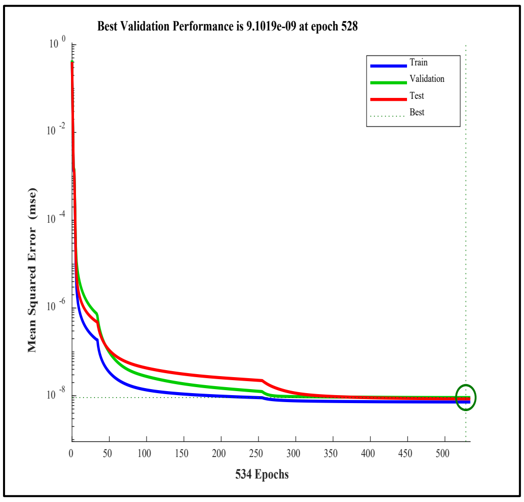


FIGURE 6: COEFFICIENT OF REGRESSION (R) PLOTS OF UNSEEN DATA NETWORKS

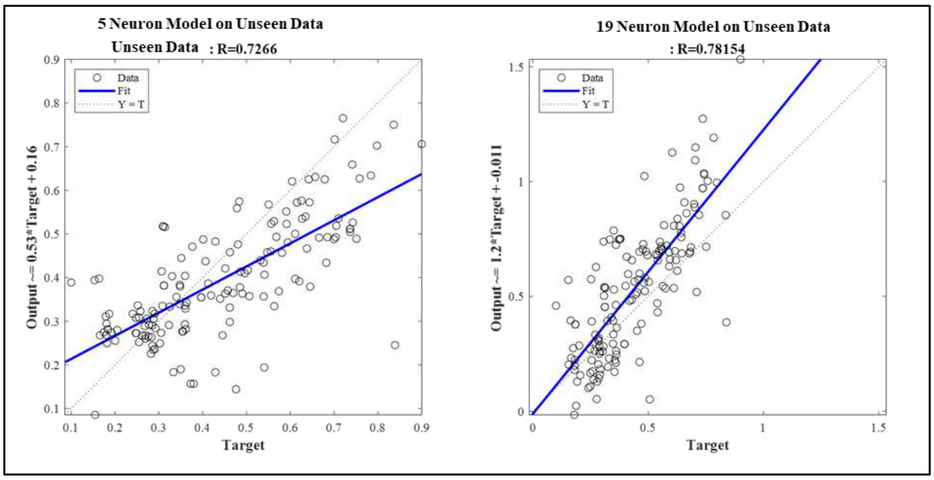


FIGURE 7: COEFFICIENT OF REGRESSION (R) PLOT OF 5 MODEL TRAINING VALIDATION TESTING.

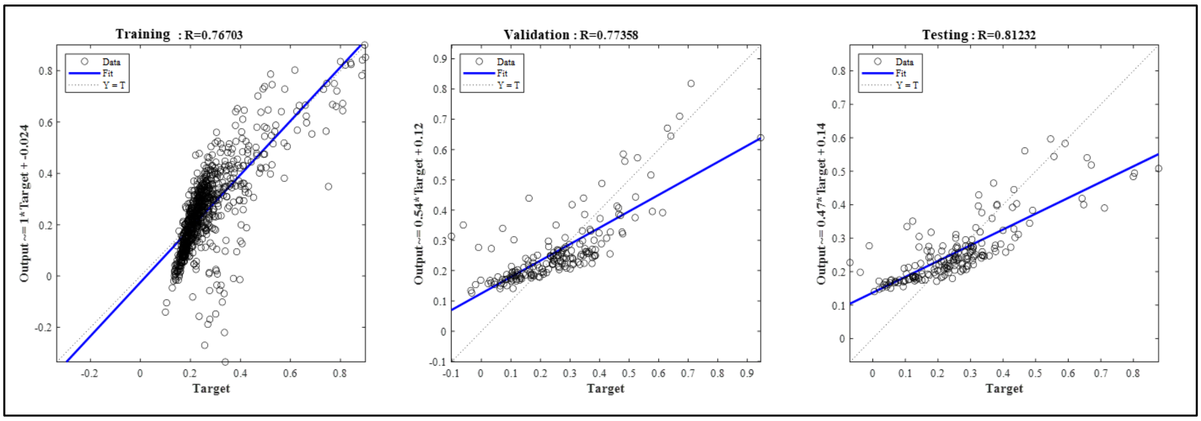


FIGURE 8: COEFFICIENT OF REGRESSION (R) PLOT OF 17 MODEL TRAINING VALIDATION TESTING.

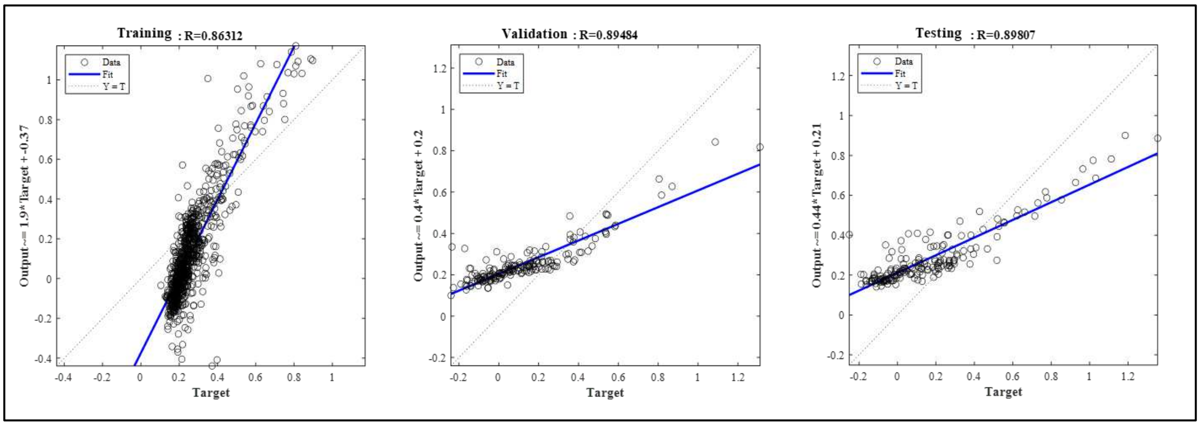
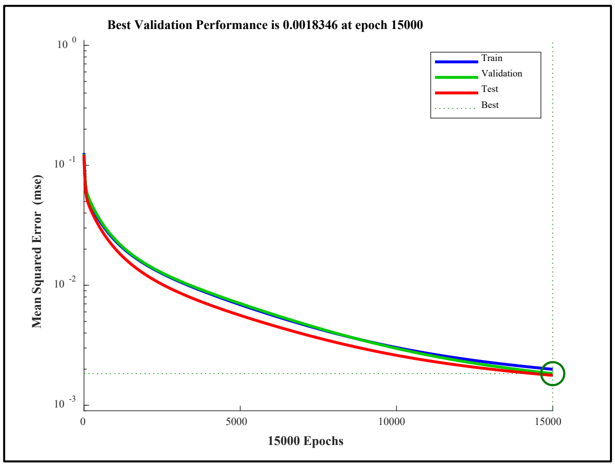


FIGURE 9: PERFORMANCE OF 5 NEURONS MODEL

FIGURE 10: PERFORMANCE OF 17 NEURONS MODEL



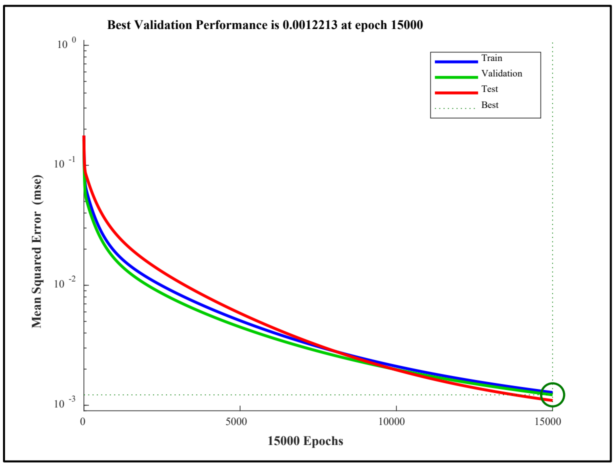


FIGURE 11: COEFFICIENT OF REGRESSION (R) PLOT ON UNSEEN DATA.

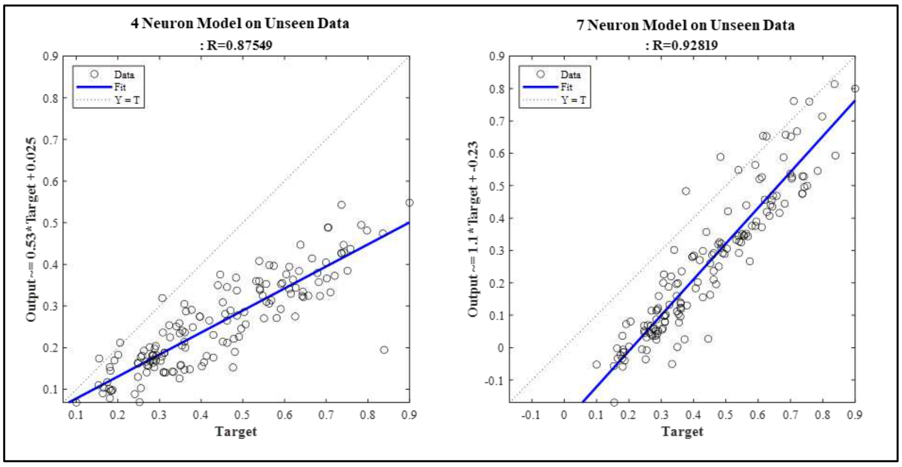


TABLE 1. DATA ACQUISITION AND STUDY AREA 2018

No.	Parameter	Max	Min	Mean	Median	SD
1	Turbidity	66.6	0.1	4.29	4	9.52
2	pH	8.99	6.58	8.08	8.10	0.35
3	Sp.Conductivity	742	8.13	308.13	311.1	89.38
4	Dissolved Solids	494	21.8	210.57	202	58.52
5	Nitrate	23.39	0.02215	3.40	3.40	3.86
6	SulphateSO4	30	0.24	6.54	6.6	3.82
7	Total hardness	246	11	123.73	128	29.10
8	Dissolved Oxygen	9.8	5	7.53	7.5	0.55
9	BOD_5	2.9	0.1	1.38	1.4	0.39
10	Total Coliform	540	1.8	45.31	43	62.84

TABLE 2. DATA ACQUISITION AND STUDY AREA 2019

No	Parameter	Max	Min	Mean	Median	SD
1	Turbidity	13.60	0.80	4.36	2.98	3.09
2	pH	8.68	7.18	7.88	7.91	0.31
3	Sp.Conductivity	416.30	170.00	285.78	274.20	49.46
4	Dissolved Solids	360.00	48.00	200.83	184.00	52.29
5	Nitrate	221.94	0.08	7.69	4.08	25.03
6	SulphateSO4	29.98	0.20	8.91	7.83	5.16
7	Total hardness	264.00	8.00	129.97	124.00	36.44
8	Dissolved Oxygen	9.00	6.90	7.92	7.85	0.49
9	BOD_5	2.80	0.10	1.29	1.20	0.42
10	Total Coliform	130.00	1.80	45.92	45.00	21.65

TABLE 8. WATER QUALITY INDEX WEIGHT CALCULATION

											Total
BIS Standards (Sn)	5	8.5	300	500	20	400	300	6	2	50	1591.5
1/Sn	0.2	0.1176	0.003333	0.002	0.05	0.0025	0.0033	0.166	0.5	0.02	1.1
K = 1/(1/ΣSn)	0.938543	0.938	0.93854	0.93854	0.9385	0.938543	0.93854	0.938	0.9385	0.9385437	9.4
W = K/Sn	0.18770	0.1104	0.003128	0.00187	0.04692	0.002346	0.003128	0.15	0.469	0.01877	1.0

Ideal Value	0	7	0	0	0	0	0	0	0		

TABLE 9. WATER QUALITY INDEX CALCULATION TABLE

No	Turbidity	pH	Sp. Conductivity	Dissolved Solids	Nitrate	Sulphate SO4	Total hardness	DO	BOD_5	Total Coliform	WQI
1	1.7	7.42	179.3	155	0.886	3.6	92	7.1	0.8	2	**45**
2	1.6	7.42	191.3	151	1.8606	3.2	88	6.9	1.2	2	**54**
3	1.5	7.45	205.6	164	0.9746	3.8	84	6.8	1.4	1.8	**58**
4	2.9	7.98	226.3	159	1.5948	3.6	80	7	1.1	1.8	**57**
5	3.4	7.98	283.7	151	2.0378	4.2	76	6.9	0.9	1.8	**54**
6	2.9	7.43	182.3	121	0.886	2.8	76	6.2	0.2	1.8	**33**
7	3.6	7.92	288.9	159	1.2404	5	88	6.6	0.8	4	**52**
8	8.9	7.86	359.1	134	0.5316	2.1	92	6.7	1.3	1.8	**83**
9	7.7	7.86	311.6	151	0.7088	2.6	88	6.6	0.8	1.8	**67**
10	4.2	7.83	215.8	139	1.0632	2.1	80	7.1	0.3	1.8	**43**
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
1191	1.3	8.62	305	230	3.544	4.3	140	7.4	1.1	48	**55**
1192	0.92	8.62	320	200	3.8984	5.3	140	7.4	1.1	49	**53**
1193	0.92	8.41	384	225	3.63703	7.12	130	6.8	1.1	40	**51**
1194	2.3	8.26	468	312	5.0059	8.48	130	6.8	1.1	47	**57**
1195	28.61	8.09	510	328	5.316	9.61	140	6.3	0.8	47	**148**
1196	8.19	7.84	192	122	4.53189	6.09	120	6	0.8	43	**70**
1197	1.1	7.84	312	218	4.53189	7.09	128	6.4	1.1	38	**51**
1198	1.1	8.22	364	219	3.544	4.4	160	6.9	1	41	**50**
1199	1.9	8.52	340	220	3.46426	4.8	156	7.8	1.2	47	**60**
1200	5.2	8.25	324.7	220	3.46426	5.56	156	7.8	1	48	**68**

TABLE 10. EVALUATING PERFORMANCE OF VARIOUS NEURAL NETWORKS (LEVENBERG)

Varying the Numbers of Neurons												Least MSE difference
Model No.	**No. of Neurons**		**Training**			**Validation**			**Testing**			
		Ref.	**MSE**	**RMSE**	**R**	**MSE**	**RMSE**	**R**	**MSE**	**RMSE**	**R**	
1	2	40	0.012	0.1097	0.4259	0.0155	0.1245	0.4855	0.168	0.1296	0.59767	0.1560
2	3	14	0.012	0.1095	0.4485	0.0165	0.1283	0.4983	0.0122	0.1104	0.5254	0.0002
3	4	48	0.0175	0.1321	0.3771	0.015	0.1224	0.5143	0.0436	0.2088	0.124	0.0261
4	5	13	0.0132	0.115	0.3243	0.0136	0.1166	0.2887	0.0109	0.1042	0.3208	-0.0023
5	6	38	0.0152	0.1231	0.3704	0.021	0.145	0.3106	0.1067	0.1292	0.36102	0.0915
6	7	28	0.107	0.1034	**0.8006**	0.0073	0.0852	0.6515	0.0118	0.1084	0.6292	-0.0952
7	8	17	0.0222	0.149	0.2634	0.0316	0.1778	0.2875	0.0198	0.146	0.4355	-0.0024
8	9	47	0.0314	0.1771	-0.0007	0.022	0.1482	0.2361	0.0308	0.1756	-0.00067	-0.0006

9	10	46	0.021	0.1451	0.2884	0.032	0.1788	0.2353	0.03	0.1732	0.0947	0.0090
10	11	19	0.0183	0.1353	0.365	0.0155	0.1243	0.3729	0.021	0.1449	0.43572	0.0027
11	12	13	0.0253	0.1591	0.6155	0.0279	0.167	0.6157	0.0245	0.1565	0.59214	-0.0008
12	13	39	0.0226	0.1503	0.6628	0.025	0.1581	0.6447	0.0221	0.1488	0.69818	-0.0005
13	14	33	0.0407	0.2018	0.0573	0.0468	0.2164	0.1255	0.0429	0.2071	0.0586	0.0022
14	15	17	0.0488	0.2209	0.5998	0.0419	0.2048	0.6442	0.053	0.2301	0.6945	0.0042
15	16	34	0.0441	0.212	0.5266	0.045	0.212	0.5156	0.0517	0.2273	0.56597	0.0076
16	17	25	0.0459	0.2142	0.2584	0.0545	0.2334	0.1723	0.0419	0.2047	0.19087	-0.0040
17	18	27	0.0458	0.2141	0.6808	0.0412	0.2029	0.6717	0.0438	0.2092	0.6795	-0.0020
18	19	31	0.0196	0.1399	0.6594	0.0167	0.1293	0.726	0.0154	0.1242	**0.74274**	-0.0042
19	20	12	0.0526	0.2293	0.3057	0.0585	0.242	0.2862	0.0594	0.2437	0.20223	0.0068
20	21	27	0.0534	0.2311	0.3479	0.0552	0.2349	0.4799	0.0645	0.2539	0.25009	0.0111
Note :	The models with lowest MSE in Training & Highest R in Testing were selected.											-0.0952
Results												
			0.0120		0.8006				**0.0109**		0.7427	

TABLE 3. PERMISSIBLE LIMITS OF THE PARAMETERS USED IN CALCULATING WQI FOR CLASS 'A'.

Parameters	**Permissible limits**
Turbidity	5
pH	8.5
Sp. Conductivity	300
Dissolved Solids	500
Nitrate mg/l	20
Sulphate (SO4) mg/l	400
Total Hardness	300
Dissolved Solids	6
Biochemical Oxygen Demand	2
Total Coliform	50

TABLE 5. PERFORMANCE OF LEVENBERG TRAINED NETWORKS ON UNSEEN DATA

R Values Levenberg-Marquardt Algorithm (Unseen Data)	
Number of Neurons	**R**
5	0.7266
19	0.7815

TABLE 4. RESULTS FROM LEVENBERG-MARQUARTH ALGORITM

R Values Levenberg-Marquardt Algorithm			
Number of Neurons	**Training**	**Validation**	**Testing**
7	0.80063	0.6515	0.6292
19	0.6594	0.7260	0.7424

FIGURE 13: COMPARISON OF ACTUAL VS PREDICTED VALUES IN TESTING PHASE USING 7 NEURON MODEL

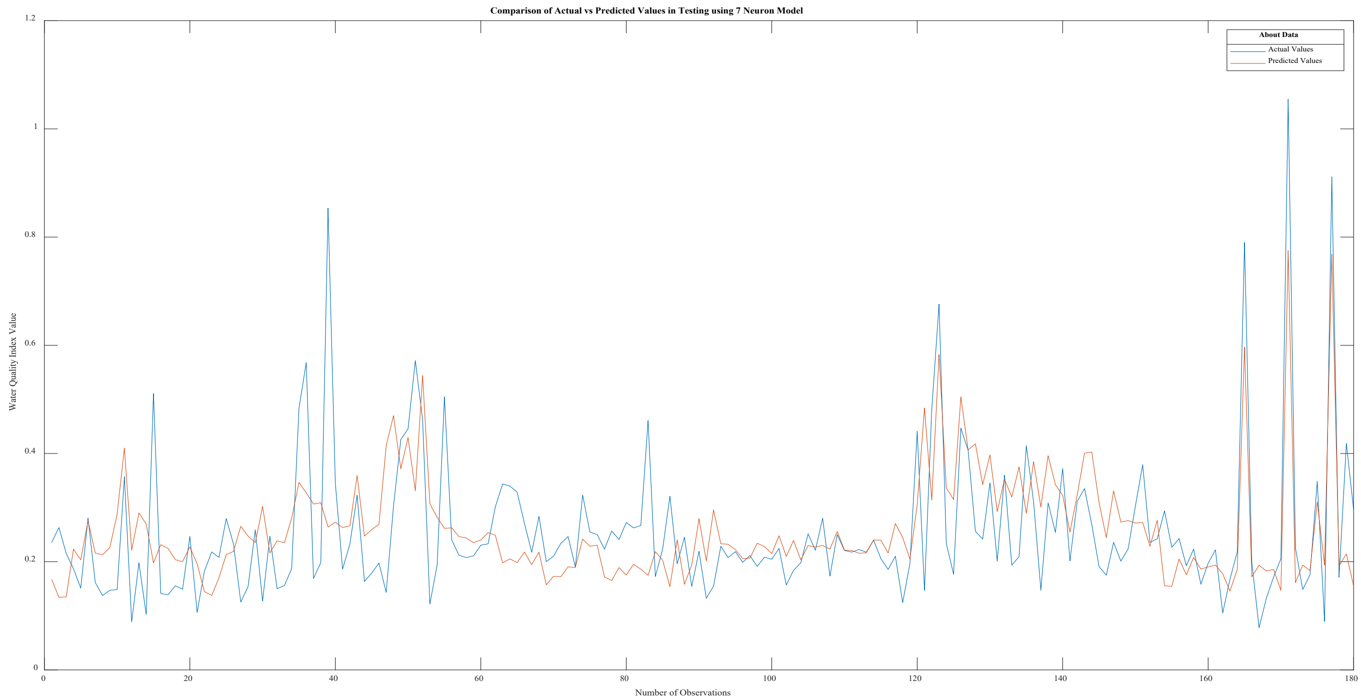


FIGURE 14: COMPARISON OF ACTUAL VS PREDICTED VALUES IN TESTING PHASE USING 19 NEURON MODEL

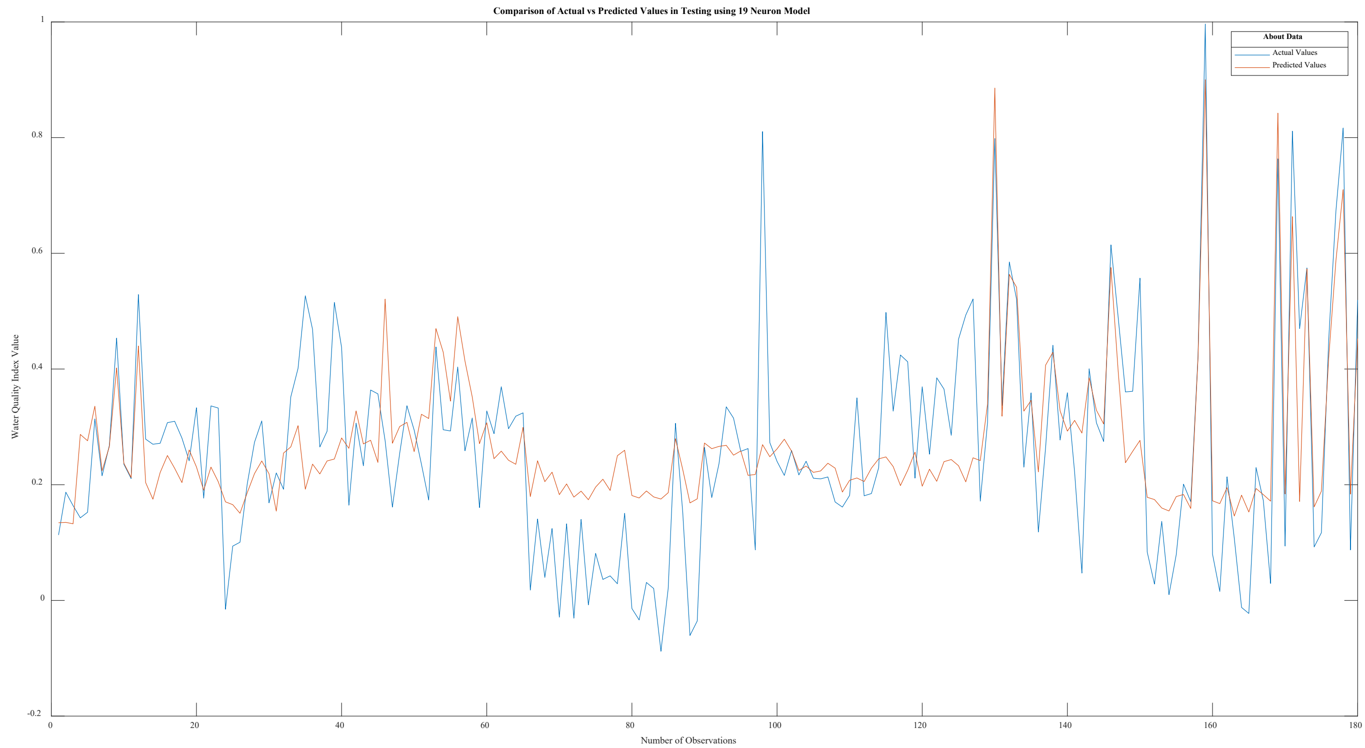


TABLE 11. EVALUATING PERFORMANCE OF VARIOUS NEURAL NETWORKS (GRADIENT DESCENT)

Varying the Numbers of Neurons												Least MSE difference
Model No.	No. of Neurons		Training			Validation			Testing			
		Ref.	MSE	RMSE	R	MSE	RMSE	R	MSE	RMSE	R	
1	2	36	0.0138	0.1175	0.6171	0.0122	0.1106	0.4872	0.0126	0.1124	0.6321	0.0012
2	3	2	0.0122	0.1102	0.5579	0.0152	0.1234	0.5738	0.0110	0.1047	0.5590	0.0012

3	4	50	0.0086	0.0929	0.6359	0.0082	0.0905	0.6119	0.0072	0.0850	0.6072	0.0014
4	5	12	**0.0015**	0.1071	0.7670	0.0104	0.1018	0.7736	0.0098	0.0990	0.8123	-0.0083
5	6	5	0.0127	0.1126	0.4068	0.0146	0.1210	0.3862	0.0104	0.1022	0.2396	0.0023
6	7	4	0.0098	0.0990	0.6897	0.0086	0.0927	0.6971	0.0086	0.0927	0.7364	0.0012
7	8	22	0.0183	0.1353	0.3552	0.0191	0.1381	0.3318	0.0200	0.1413	0.2398	-0.0017
8	9	13	0.0099	0.0996	0.7280	0.0118	0.1085	0.6892	0.0113	0.1061	0.7661	-0.0014
9	10	37	0.0255	0.1596	-0.3429	0.0336	0.1824	- 0.4060	0.0231	0.1520	-0.4872	0.0024
10	11	19	0.2480	0.1576	0.6618	0.0333	0.1825	0.6011	0.0302	0.1736	0.5911	0.2178
11	12	28	0.0165	0.1283	0.3501	0.0199	0.1409	0.2858	0.0167	0.1292	0.3835	-0.0002
12	13	19	0.0402	0.2004	0.3121	0.0421	0.2052	0.5044	0.0347	0.1863	0.3150	0.0055
13	14	39	0.0361	0.1910	0.5853	0.0350	0.1870	0.5082	0.0425	0.2062	0.5079	-0.0064
14	15	27	0.0436	0.2089	-0.0209	0.0573	0.2395	- 0.2091	0.0476	0.2181	0.0115	-0.0040
15	16	33	0.0205	0.1433	0.5802	0.0214	0.1463	0.5807	0.0178	0.1334	0.6845	0.0027
16	17	18	**0.0440**	0.2098	**0.8631**	0.0399	0.1998	0.8948	0.0435	0.2085	**0.8981**	**0.0005**
17	18	5	0.0319	0.1785	0.7064	0.0323	0.1797	0.7388	0.0248	0.1574	0.7751	0.0071
18	19	42	0.0489	0.2212	0.2439	0.0524	0.2290	0.3158	0.0511	0.2261	0.24762	-0.0022
19	20	27	0.0465	0.2155	0.3537	0.0464	0.2154	0.4508	0.0505	0.2247	0.2944	-0.0040
20	21	16	0.0539	0.2323	0.2982	0.0461	0.2147	0.3123	0.0444	0.2108	0.4149	0.0095
Note :	The models with lowest MSE in Training & Highest R in Testing were selected.											-0.0083
Results			0.0015		0.8631						0.8981	

TABLE 7. PERFORMANCE OF GRADIENT DESCENT TRAINED NETWORKS ON UNSEEN DATA

R Values Gradient Descent (Unseen Data)	
Number of Neurons	**R**
4	0.8755
7	0.9282

TABLE 6. RESULTS FROM GRADIENT DESCENT ALGORITHM

R Values Gradient Descent			
Number of Neurons	**Training**	**Validation**	**Testing**
5	0.7670	0.7736	0.8123
17	0.8631	0.8948	0.8981

FIGURE 15: COMPARISON OF ACTUAL VS PREDICTED VALUES IN TESTING PHASE USING 5 NEURON MODEL

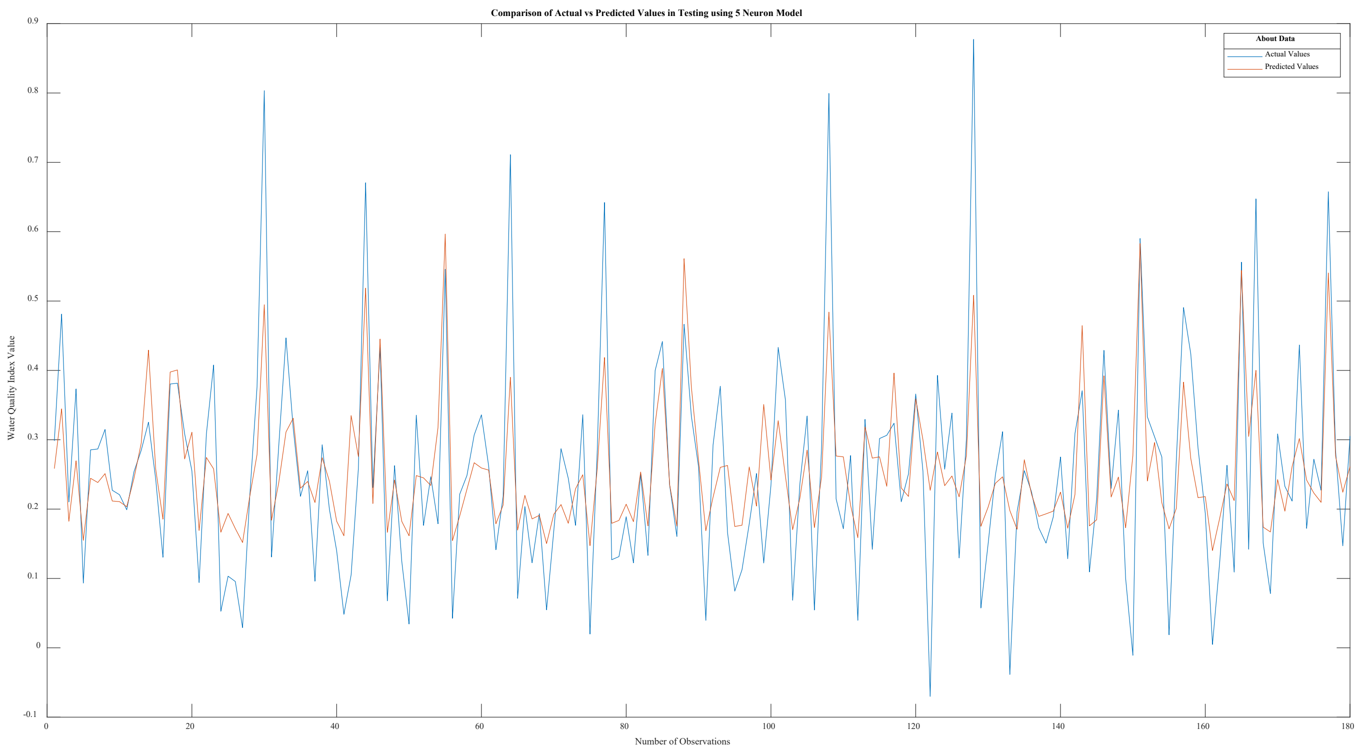
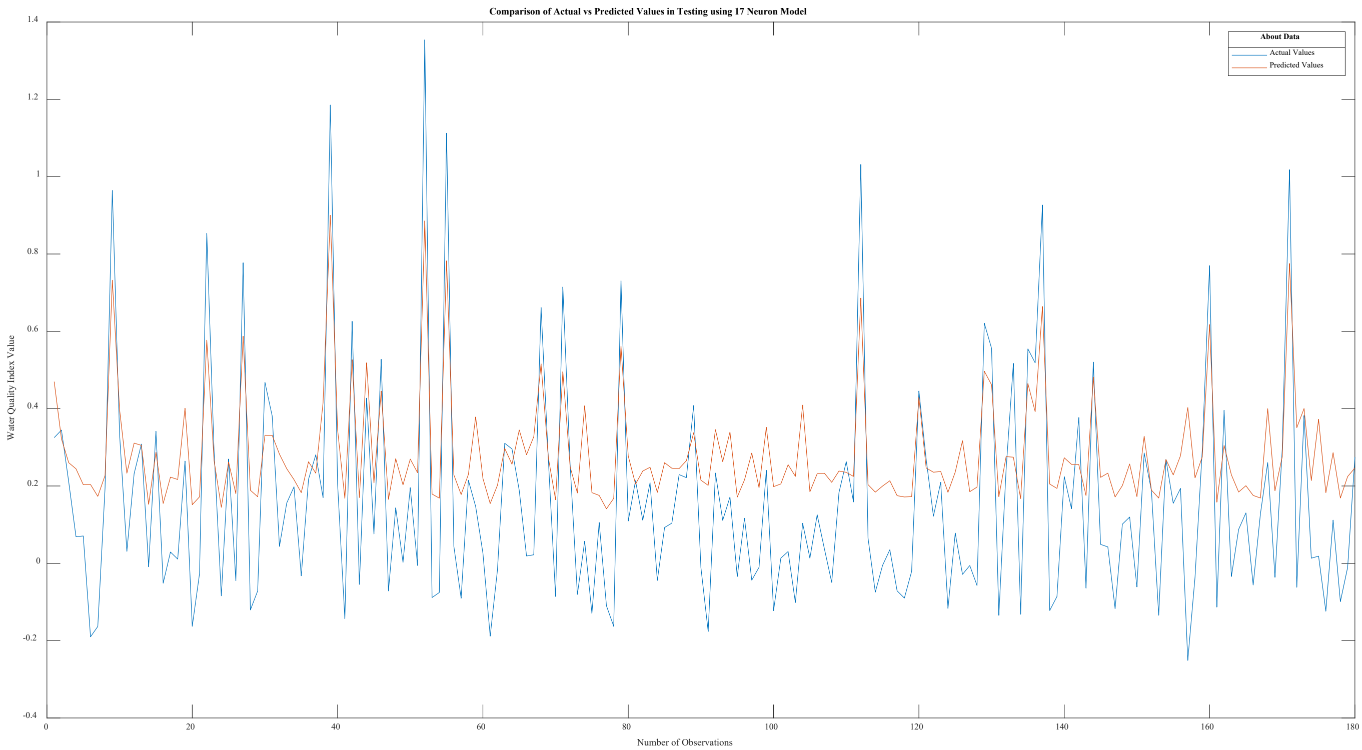


FIGURE 16: COMPARISON OF ACTUAL VS PREDICTED VALUES IN TESTING PHASE USING 17 NEURON MODEL



3.2 Discussion

The Levenberg algorithm and Gradient Descent algorithm both performed significantly well in this study. But when compared the results, it is clear that Gradient descent performed slightly better than Levenberg. The number of neurons in the hidden layers is the second important factor here, and both the algorithms performed somewhat average in this regard. The best results from the

Levenberg algorithm were observed with 7 & 19 neurons in the hidden layer. And the best result from the Gradient descent algorithm was observed with 5 & 17 neurons in the hidden layer.

In Madhya Pradesh, there are 10 river basins, but there is no central infrastructure that would help us understand the current state of the river and predict what are the areas which need the most attention. In this paper, an attempt has been made to resolve this issue by developing a model for the prediction of the water quality of river Narmada which is one of the most import rivers in MP. Artificial neural networks are very powerful in handling complex problems like this and observing patterns in data. The results can be improved by performing sensitivity analysis & Hyperparameter selection which is a part of another study of this topic. Further, adding more data & performing time series analysis, will make the results much more efficient and accurate. The follow-up study will primarily be based on improving results by various methods of parameter selection and adding data. In this study, only a single hidden layer was used, and the number of neurons was varied, in the future study it will be also be observed how deep neural networks (artificial neural networks with more than 1 hidden layer) performs with the improved dataset.

IV. ACKNOWLEDGMENT

I would like to thank, my guide Dr. Sunayana, Scientist, CSIR-NEERI, for guiding me through every point during this study. I would also like to thank Prof. D.C. Rahi, Department of Civil Engineering, Jabalpur Engineering college for his insights on this work.

V. DECLARATION

I *Shubham Lakhera* solemnly declare that this research, "DEVELOPMENT OF WATER QUALITY PREDICTION MODEL FOR NARMADA RIVER USING ARTIFICIAL NEURAL NETWORKS" is based on my own work carried out during the March 2021 to November 2021, under the supervision of Dr. Sunayana & D.C Rahi. I assert the statements made and conclusions drawn are an outcome of my research work.

VI. ETHICAL APPROVAL

Not Applicable.

VII. CONSENT TO PARTICIPATE

Both Authors has been given consent to participate and accepted this research.

VIII. CONSENT TO PUBLISH

Both Authors have given their consent for this publication.

IX. AUTHORS CONTRIBUTIONS

The Writing of Code, Data Collection, Processing, Analysis of Results, & Rest of the work is done by Shubham Lakhera. It is guided by Dr. Sunayana (Scientist, CSIR-NEERI). Occasionally, the progress of work has been reviewed and suggestions were made by D.C Rahi, (Assistant Professor, Jabalpur Engineering College).

X. FUNDING

No funding was received to assist with the preparation of this manuscript.

XI. COMPETING INTERESTS

There are no self-conflict of interest authors between the authors of the paper.

XII. AVAILABILITY OF DATA AND MATERIALS

Not Applicable

REFERENCES

Ahuja, Satinder, Monitoring water quality: Pollution assessment, analysis, and remediation. Newnes, 2013.

Aldhyani, T. H. H., Al-Yaari, M., Alkahtani, H., & Maashi, M. (2020). Water Quality Prediction Using Artificial Intelligence Algorithms. Applied Bionics and Biomechanics, 2020.

Al-Adhaileh, M. H., & Alsaade, F. W. (2021). Modelling and prediction of water quality by using artificial intelligence. Sustainability (Switzerland), 13(8).

Brown, R. M., McClelland, N. I., Deininger, R. A., & Tozer, R. G. (1970). A water quality index-do we dare. Water and sewage work, 117(10).

Boger, Z., & Guterman, H. (1997, October). Knowledge extraction from artificial neural network models. In 1997 IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation (Vol. 4, pp. 3030-3035). IEEE.

Blum, Adam. Neural networks in C++ an object-oriented framework for building connectionist systems. John Wiley & Sons, Inc., 1992: P.60

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Computer Science, 7, 1–24.

Dawood, T., Elwakil, E., Novoa, H. M., & Gárate Delgado, J. F. (2021). Toward urban sustainability and clean potable water: Prediction of water quality via artificial neural networks. Journal of Cleaner Production, 291.

Horton, Robert K. “An index number system for rating water quality.” J Water Pollut Control Fed 37.3 (1965): 300-306.

Jain, A., Nandakumar, K., & Ross, A. (2005). Score normalization in multimodal biometric systems. Pattern Recognition, 38(12), 2270–2285.

Y. Khan and C. S. See, "Predicting and analyzing water quality using Machine Learning: A comprehensive model," 2016 IEEE Long Island Systems, Applications and Technology Conference (LISAT), 2016, pp. 1-6.

Kim, P. (2017). MATLAB Deep Learning. In MATLAB Deep Learning. Apress.

Leong, W. C., Bahadori, A., Zhang, J., & Ahmad, Z. (2021). Prediction of water quality index (WQI) using support vector machine (SVM) and least square-support vector machine (LS-SVM). International Journal of River Basin Management, 19(2), 149–156.

Levenberg1, K., & Arsenal, F. (n.d.). A method for the solution of certain non-linear problems in least squares.

Linoff, Gordon S., and Michael JA Berry. Data mining techniques: for marketing, sales, and customer relationship management. John Wiley & Sons, 2011. P. 323

Najah, A., El-Shafie, A., Karim, O. A., & El-Shafie, A. H. (2013). Application of artificial neural networks for water quality prediction. Neural Computing and Applications, 22, 187-201.

Nayak, J. G., Patil, L. G., & Patki, V. K. (2021). Artificial neural network-based water quality index (WQI) for river Godavari (India). Materials Today: Proceedings.

National Mission for Clean Ganga, Ministry of Jal Shakti by Department of water resources, river development & ganga rejuvenation, Government of India, Page: 120-121.

Neural Networks - A Comprehensive Foundation - Simon Haykin. (n.d.).

Niles, Henry E. "Correlation, causation and Wright's theory of" path coefficients"." Genetics 7.3 (1922): 258.

Nilsson, N. J. (1996). Introduction to machine learning.

Nouraki, A., Alavi, M., Golabi, M., & Albaji, M. (2021). Prediction of water quality parameters using machine learning models: a case study of the Karun River, Iran. Environmental Science and Pollution Research, 28(40), 57060–57072.

Salari, M., Teymouri, E., & Nassaj, Z. (n.d.). Application of an Artificial Neural Network Model for Estimating Water Quality Parameters in the Karun River, Iran. Journal of Environmental Treatment Techniques 2021, 9(4), 720–727.

Satinder Ahuja - Handbook of Water Purity and Quality-Academic Press (2009). (n.d.).

Surface Water Quality Criteria for different uses (Specified by CPCB, 1979 and the Bureau of Indian standards, 1982)

Surface Water Quality Standards as per IS: 2296 Class A, Chapter 9 “Relevant Indian Standards”

Tripathi, M., & Singal, S. K. (2019). Use of Principal Component Analysis for parameter selection for development of a novel Water Quality Index: A case study of river Ganga India. Ecological Indicators, 96, 430–436.

Uddin, M. G., Nash, S., & Olbert, A. I. (2021). A review of water quality index models and their use for assessing surface water quality. In Ecological Indicators (Vol. 122). Elsevier B.V.

Vijay, S., & Kamaraj, K. (2021). Prediction of Water Quality Index in Drinking Water Distribution System Using Activation Functions Based Ann. Water Resources Management, 35(2), 535–553.

Wagh, V. M., Panaskar, D. B., Muley, A. A., Mukate, S. V., Lolage, Y. P., & Aamalawar, M. L. (2016). Prediction of groundwater suitability for irrigation using artificial neural network model: a case study of Nanded tehsil, Maharashtra, India. Modeling Earth Systems and Environment, 2(4).

Wechmongkhonkon, S., N. Poomtong, and S. Areerachakul. "Application of artificial neural network to classification surface water quality." World Academy of Science, Engineering and Technology 6.9 (2012): 574-578.

Willmott, Cort J., and Kenji Matsuura. "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance." Climate research 30.1 (2005): 79-82

Y. Xiang and L. Jiang, "Water Quality Prediction Using LS-SVM and Particle Swarm Optimization," 2009 Second International Workshop on Knowledge Discovery and Data Mining, 2009, pp. 900-904.