

Article

Neural Network-Based Modeling of Water Quality in Jodhpur, India

Krishna Kumar Sinha ^{1,*}, Mukesh Kumar Gupta ^{1,*}, Malay Kumar Banerjee ¹, Gowhar Meraj ^{2,3}, Suraj Kumar Singh ⁴, Shruti Kanga ³, Majid Farooq ^{2,3}, Pankaj Kumar ^{5,*} and Netrananda Sahu ⁶

¹ Department of Energy and Electrical Engineering, Suresh Gyan Vihar University, Jaipur 302017, Rajasthan, India; krishna.60600@mygyanvihar.com (K.K.S.); malay.banerjee@mygyanvihar.com (M.K.B.)

² Government of Jammu & Kashmir, Department of Ecology, Environment & Remote Sensing, SDA Colony Bemina, Bemina 190018, Srinagar, India; gowharmeraj@gmail.com (G.M.); majid_rsgis@yahoo.com (M.F.)

³ Centre for Climate Change, Suresh Gyan Vihar University, Jaipur 302017, Rajasthan, India; shruti.kanga@mygyanvihar.com

⁴ Centre for Sustainable Development, Suresh Gyan Vihar University, Jaipur 302017, Rajasthan, India; suraj.kumar@mygyanvihar.com

⁵ Institute for Global Environmental Strategies, Hayama 240-0115, Kanagawa, Japan

⁶ Department of Geography, Delhi School of Economics, University of Delhi, Delhi 110007, Delhi, India; nsahu@geography.du.ac.in

* Correspondence: mukeshkr.gupta@mygyanvihar.com (M.K.G.); kumar@iges.or.jp (P.K.)



Citation: Sinha, K.K.; Gupta, M.K.; Banerjee, M.K.; Meraj, G.; Singh, S.K.; Kanga, S.; Farooq, M.; Kumar, P.; Sahu, N. Neural Network-Based Modeling of Water Quality in Jodhpur, India. *Hydrology* **2022**, *9*, 92. <https://doi.org/10.3390/hydrology9050092>

Academic Editors: Rafik Absi, Carlo Gualtieri, Qi Meilan and Kwok-Wing Chau

Received: 30 March 2022

Accepted: 18 May 2022

Published: 20 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: In this paper, the quality of a source of drinking water is assessed by measuring eight water quality (WQ) parameters using 710 samples collected from a water-stressed region of India, Jodhpur Rajasthan. The entire sample was divided into ten groups representing different geographic locations. Using American Public Health Association (APHA) specified methodology, eight WQ parameters, viz., pH, total dissolved solids (TDS), total alkalinity (TA), total hardness (TH), calcium hardness (Ca-H), residual chlorine, nitrate (as NO_3^-), and chloride (Cl^-), were selected for describing the water quality for potability use. The quality of each parameter is examined as a function of the zone. Taking the average parametric values of different zones, a unique number was used to describe the overall quality of water. It was found that the average value of each parameter varies significantly with zones. Further, we used neural network (NN) modeling to map the nonlinear relationship between the above eight parametric inputs and the water quality index as the output. It can be observed that the NN designed in the present work acquired sufficient learning and can be satisfactorily used to predict the relational pattern between the input and the output. It can further be observed that the water quality index (WQI) from this work is highly efficient for a successful assessment of water quality in the study area. The major challenge to uniquely describing the drinking water quality lies in understanding the cumulative effect of various parameters affecting the quality of water; the quantified figure is subjected to debate, and this paper addresses the difficulty through a novel approach. The framework presented in this work can be automated with appropriate equipment and shall help government agencies understand changing water quality for better management.

Keywords: water quality parameters; BIS standards; water quality index; neural network

1. Introduction

Water is the largest consumable of human beings and is the primary requirement for the sustenance of animals and plants. The water quality of rivers and lakes is of extreme importance as it impacts the well-being of humans and the sustenance of aquatic animals. Industrial wastewater-led contamination of source water is one of the significant potential threats to water quality, particularly in urban areas [1]. Moreover, water quality in these water bodies is the primary determinant of ecological balance within the living world [2]. In

contrast, the continued enhancement in water consumption from natural sources has led to its quality degradation [3]. It is known that the characteristic quality of water required for drinking, industrial activities, or agriculture varies in terms of the concentration of various quality parameters; thus, different human activities need different parametric values of quality variables. Industrialization and indiscriminate use of chemical fertilizers/pesticides in agricultural activities have been continuously polluting the aquatic environment, leading to dwindling aquatic biota. Moreover, human beings are getting infected with water-borne diseases by using contaminated water [4].

Several technical parameters decide if water is suitable for a specific activity or not. The quality of drinking water is a prime concern in today's scenario. As mentioned earlier, the ever-increasing global population, the significant enhancement in the use of water for various purposes, and the generation of wastewater by default signal that the whole world is to pass through a tremendous water crisis unless some innovative technologies are invoked [5]. Most countries worldwide have designed innovative ways to minimize water pollution in natural resources and evolve new technologies for water purification. The flows of many natural resources are transnational, so the mitigation strategy must be cooperative and inclusive. With such a realization, the present-day drive for preserving water resources on Earth has become an essential scientific and technological activity [6].

Water's physical, chemical, and biological characteristics are used to describe water quality for drinking, sustaining aquatic species, and various industrial and agricultural activities [7]. Various studies have demonstrated that parameters such as biological oxygen demand (BOD), chemical oxygen demand (COD), dissolved oxygen, pH, total dissolved solids (TDS), total alkalinity (TA), total hardness (TH), calcium hardness (Ca-H), residual chlorine, nitrate (as NO_3^-), and chloride (Cl^-) must have their values within specific ranges to determine the suitability of the water usage [8,9]. Some parameters are likely highly satisfactory in a water sample, whereas others are not. Taking such a variation into consideration, it has been advocated by several researchers that there should be a unique identity to represent the overall quality of water. This is done by describing the water quality index (WQI) by differently derived numbers [10,11]. The water quality index, whichever way it might have been defined, aims to help people understand the suitability of given water for a specific use. The water quality index (WQI) developed by the Canadian Council of Ministers of the Environment (CCME) has been employed to judge the suitability of water at several geographical locations, including a few in the case of lakes situated in India [12]. While a two-layer time-variable model has been developed to quantify seasonal variations of pH and alkalinity levels [13] for specific cases, there are recorded efforts to research water quality planning and management [14]. Mathematical modeling and computer simulation for predicting the degradation of water quality as a function of time and space, including many statistical modeling techniques, have been employed to estimate pollution loadings in water. It is defined in terms of the water quality index, and a decision support system is embedded in the analyses to decide on the efficacy of the probable solution.

Among the different purposes for which water quality indices must be used, the quality assessment of drinking water by suitable indexing is of paramount importance. The World Health Organization (WHO) and the Bureau of Indian Standards (BIS) have determined the desirability ranges of various water quality parameters to make up potable water. As mentioned earlier, industrial wastewater discharge contaminates the water sources that need to be appropriately treated for further use. This necessitates the development of a new but simple protocol to determine the water quality index for drinking purposes. Using specifications set by various agencies, it seems possible to set the acceptable limits (both upper and lower) of water quality parameters that are harmonious with the available parametric range within the study area, that is, the Jodhpur district of India. The present study to design a water quality index in a more straightforward but pragmatic way is not aimed to comply with regulations or to violate specifications. Many workers have integrated the quality parameters in a single number, i.e., the water quality index (WQI).

However, these approaches do not reveal the inherent input–output relationship, which is mostly nonlinear. Therefore, an attempt is made to get rid of this difficulty by mapping the nonlinearity in the relationship of parametric quality contribution to the overall quality index. The difficulty has been capturing a nonlinear relationship between a set of input and output and recognizing relational patterns between them using computational techniques, which has not been abundantly documented in the literature.

Realizing the importance of monitoring and predicting the changing water quality, many researchers have focused attention on modeling water quality as a function of several variables [15]. A report on artificial neural network (ANN) modeling [16] employed both multilayer perceptron learning and a neural network-based radial basis function (RBF). Reportedly, a predictive capacity has enabled very effective water resources management in South Africa. Moreover, the ANN has been used elsewhere to predict the water quality parameters for a period of one year, such that better control over the water quality for irrigation purposes may be ensured. The ANN modeling used for this purpose was demonstrated to have satisfactorily predicted water quality [17].

The nonstationary character of coastal water is a critical problem so far as its assessment in respect to space and time is concerned; such a nonlinear system is reported to have been ably modeled by researchers by proposing what is called a geographically neural network weighted regression (GNNWR) model. This can predict a realistic water quality distribution over the entire region of study [18]. An ANN was also used to create a water quality index by training the network by using five important and universally acceptable water quality parameters. Although the experiment was carried out with water from the Indian subcontinent, the result can be used globally [19]. It was demonstrated that an increasing number of members in the training dataset could enhance the regression value at an increased learning rate. A five-layer network produced the best result [19]. Elegant research on implementing artificial intelligence algorithms for predicting water quality index is documented in the literature [20]. An artificial neural network model and the development of a short-term memory deep learning algorithm were reported. Moreover, the same group reported using three machine learning algorithms, viz., support vector machine, naïve Bayes, and K-nearest neighbor (KNN). The models worked fine. Since ANN has proven pattern recognition capability, early workers attempted to examine if it could classify water quality parameters. The measurement data of water quality parameters, viz., pH and dissolved oxygen, were used for training and testing; this finally yielded an 80% accuracy in classifying quality parameter data at a 0.468 root mean square error (RMSE) [21].

As may be noted from the foregoing discussions, many other techniques can be applied for learning, and a large number of approximators can also be used; however, one needs to write new code and they should be capable of interfacing seamlessly with MATLAB. Therefore, as an exploratory work, it seems simple and logical to make use of neural network modeling in the present case to map the relation between common drinking water quality parameters with a well-defined water quality index. This has compatibility with the claims of the previous authors as mentioned above. The available information and the need to understand the impact of the changing values of water quality parameters on water potability enticed the authors to conduct the water quality modeling with an artificial neural network-based approach. Hence, neural network modeling is carried out with the measured drinking water data from a specific region in the present investigation. For this study, the water quality parameters estimated at different locations in Jodhpur, a water-stressed district of India, were chosen for neural network modeling.

2. Materials and Methods

In the present investigation, drinking water quality assessment and predictions were conducted in India's water-stressed state, Rajasthan. Figure 1a shows the location of the state of Rajasthan on the map of India; the annexed diagram shows the Jodhpur district, which is known to suffer from water scarcity, and water samples for the study have been

collected from the Mandore Block of Jodhpur district. Figure 1b shows the map of the Jodhpur district, pointing out the location of Mandore, from which 710 water samples were collected for testing. Figure 1c shows the Mandore area; wherein points mark the sampling sites. Moreover, to provide the geographic location of the sampling, the latitude and longitude of a few sites at the edges of the Mandore are also shown in Figure 1c. The dataset collected from 710 locations contained eight significant parameters, as shown in Tables 1 and 2.

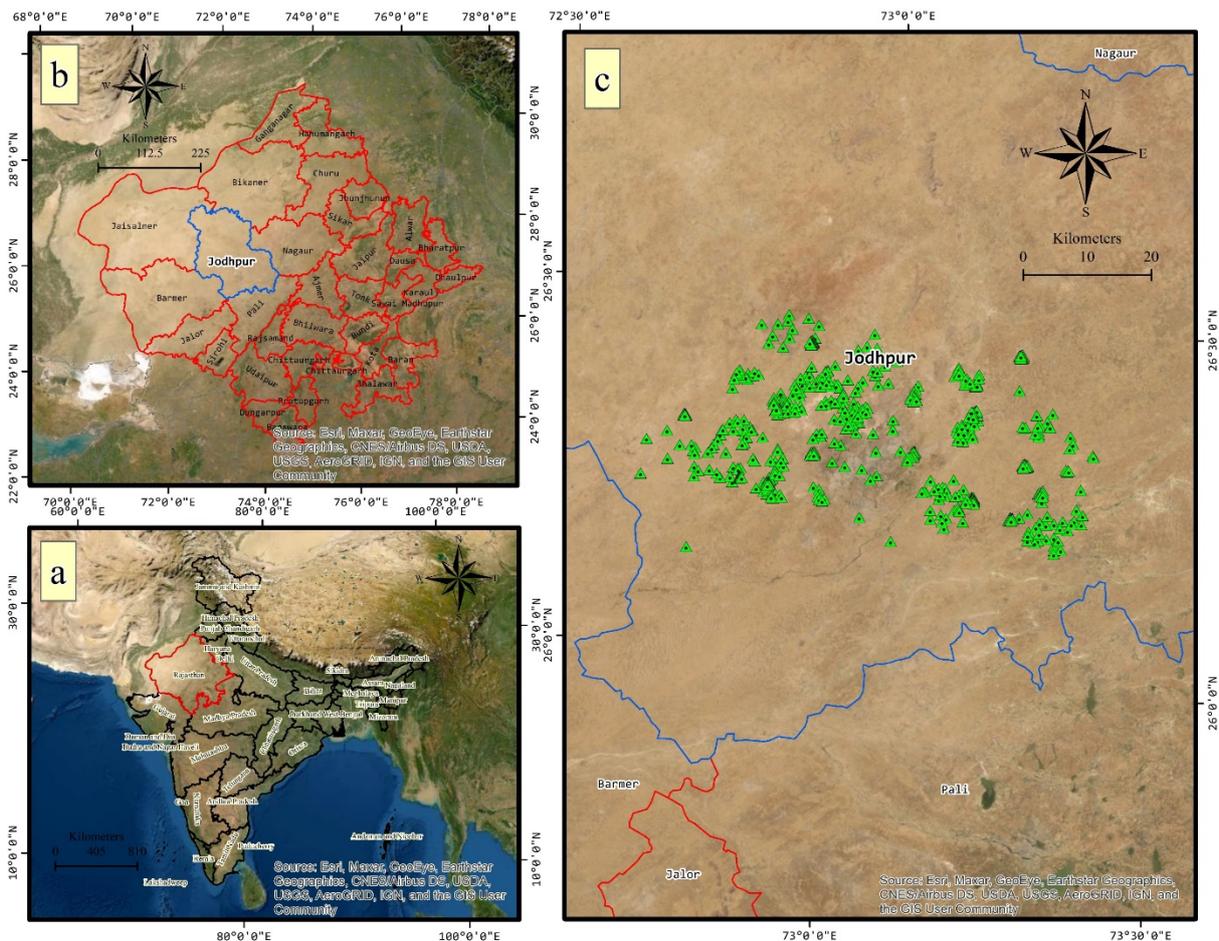


Figure 1. (a) Location of Rajasthan with respect to the union of India. (b) Location of Jodhpur district with respect to the State of Rajasthan. (c) Locations of drinking water samples in the Jodhpur district (Mandore area) with respect to Jodhpur.

Table 1. Allowable range of drinking water quality parameters in line with BIS standards [22].

Parameters	Lower Limit	Upper Limit
pH	7.0	8.5
Alkalinity (as CaCO ₃ , mg/L)	200	600
Hardness (as CaCO ₃ , mg/L)	200	600
Chloride (mg/L)	250	1000
Nitrate (mg/L)	0	45
Fluoride (mg/L)	1	1.5
TDS (mg/L)	500	2000
Residual Chlorine (mg/L)	0.2	0.5

Table 2. Average values of water quality parameters in the Jodhpur region in Rajasthan.

Zone	pH	TA (mg/L)	TH (mg/L)	Cl (mg/L)	Nitrate (mg/L)	Fluoride (mg/L)	TDS (mg/L)	Residual Chlorine (mg/L)
1	7.8 (0.44)	283.8 (116.9)	337.8 (164.9)	518.3 (185.8)	15.8 (7.8)	0.61 (0.39)	1484.8 (501.1)	0.085 (0.099)
2	7.7 (0.36)	312.7 (134.8)	316.1 (134.9)	532.8 (153.1)	15.4 (10.4)	0.82 (0.48)	1469.4 (388.7)	0.028 (0.07)
3	7.8 (0.42)	273.8 (138.7)	259.3 (144)	494.8 (231)	16.9 (7.8)	0.65 (0.44)	1493.8 (447.5)	0.056 (0.09)
4	7.8 (0.36)	290.4 (133.8)	372.8 (159.4)	500.4 (229.1)	23.4 (12.5)	0.85 (0.47)	1622.5 (664.7)	0.02 (0.06)
5	7.7 (0.29)	274.9 (120.8)	335.4 (163.7)	517.5 (274.9)	24.0 (10.7)	0.86 (0.52)	1551.4 (596.7)	0.01 (0.046)
6	7.7 (0.25)	236.1 (126.9)	289.4 (143.6)	464.5 (281.4)	21.9 (12.1)	0.81 (0.49)	1347.0 (731.6)	0.02 (0.063)
7	7.9 (0.37)	256.3 (109.8)	361.9 (187.6)	619.2 (335.5)	17.9 (13.4)	0.78 (0.42)	1666.6 (732.6)	0 (0)
8	7.8 (0.24)	262.3 (108.4)	336.5 (173)	659.7 (332.5)	21.6 (9.7)	0.87 (0.52)	1728.3 (690.9)	0.01 (0.033)
9	7.8 (0.27)	253.5 (101.9)	364.2 (159.4)	675.8 (380.9)	22.1 (11.4)	0.86 (0.51)	1711.1 (719.9)	0.01 (0.033)
10	7.8 (0.24)	254.8 (120.9)	316.6 (175.4)	632.9 (363.9)	19.9 (11.8)	0.67 (0.43)	1475.3 (601.8)	0.03 (0.67)
Average value of all zones	7.8	269.9	329.0	561.6	19.8	0.78	1555.0	0.03

Numbers in brackets denote the standard deviation of measured values in a zone.

Divided into ten zones, 710 water samples were collected from the Jodhpur region of Rajasthan India. A total of 71 water samples were collected from each zone. Water quality parameters were measured by following the standard procedure. Polypropylene bottles (1 L in size) were used to collect the drinking water samples. All the bottles were rinsed with dilute acid before being cleaned with distilled water. These cleaned bottles were used to collect the test samples. Finally, the bottles were rinsed thrice with the water samples to be collected. These were dried in the oven before collecting the test samples. The bottles containing the water samples were stored in a refrigerator until the water samples were subjected to analysis.

The water quality parameters, such as pH, TDS, TA, TH, fluoride, NO₃, Cl⁻, and residual chlorine, were tested as per the standard methods prescribed by APHA. While there could be other specifications, we have followed the techniques specified by APHA due to their universal acceptability. For water quality assessment, the measured test values were evaluated against the recommended standard as per BIS, as presented in Table 1. For ease in modeling, we took the permissible limit in BIS standard as the guidelines and used these values as the acceptable lower limit of a parameter, wherever applicable. The average test values for each of the 71 samples of the individual zones were determined and are presented in Table 2. Moreover, Table 2 shows the standard deviation for each parameter for each zone. The same table also presents the global average values of the individual quality parameters, which is the average of the parametric standards from each zone. Assuming average values of each parameter as ten inputs, a quality index algorithm was designed to check the quality of water.

3. Results and Discussion

3.1. Water Quality Evaluation

Figure 2a–h shows the variation of average values of various water quality parameters as a function of locations; the zones were selected arbitrarily and did not represent any well-defined functional relation with distance or other geographic or demographic parameters. Thus, the results shown in Figure 2a–h indicate that the quality parameters vary from place to place. It can be observed that there is considerable variation in the magnitude of average values of each quality parameter across the selected zones. In Figure 2a, it is revealed that the pH value is different in different zones, and its value lies within a range from 7.70 to 7.88 and hence is higher than the most desirable value of pH, which is 7. Similarly, another parameter, viz., total alkalinity (TA), is seen to have varied from

236.06 mg/L to 312.68 mg/L as shown in Figure 2b. The range of variation of the average values of the other water quality parameters can be seen to vary within the ranges, such as TH (259.30 mg/L to 372.82 mg/L), chloride (464.507 mg/L to 675.775 mg/L), nitrate (15.39 mg/L to 24.01 mg/L), fluoride (0.61 mg/L to 0.87 mg/L), TDS (1347.04 mg/L to 1728.31 mg/L), and residual chlorine (0–0.01); the variations are shown in Figure 2c–h, respectively. It is important to note that the TDS of water in Jodhpur is much higher than the desired value, ~less than 500 mg/L. However, in water-stressed regions, the higher value (2000 mg/L) is accepted as per the BIS standard; suitable TDS removal techniques may be adopted in such cases. It should be reiterated that ideally the TDS must be kept below 500 mg/L; although in a water-stressed region such as the Jodhpur district of Rajasthan, India, one may make use of higher TDS water. However, TDS removal techniques are quite simple, especially with solar energy. Normally, hotter areas are water-stressed (e.g., California, USA) and in most cases, one would come across high TDS water; to make water abundantly available for drinking; the simpler solar heating technique may be applied, or even boiling the water will reduce the TDS level. Considering this, we took 500 mg/L as the lower limit and 2000 mg/L as the higher limit of acceptance. Admittedly, a TDS value less than the permissible value as per the BIS standard (500 mg/L) is always better. In the present case, we took 500 mg/L to be the most desirable value which can be comfortably set as the lower limit of acceptance without regard to the beneficial consequences for the case of still lower TDS values. Moreover, it was also recognized that a zero TDS value must not index the best quality with respect to the TDS value.

Similarly, a threshold value of 0.2 was set as the acceptable lower limit for chlorine. In cases where its value goes below this threshold, it may present a health concern. Similarly, for fluoride, the lower limit was taken to be 1, notwithstanding if a lower fluoride level is good for health. Noting that the average of 71 samples from a zone varied with the location of the zones, it seemed wise to examine if the average of the zonal averages of each parameter can bear significance in determining the overall quality of drinking water for the entire region.

The overall average value of each water quality parameter for the entire region was obtained by taking the average of the individual quality average for each zone; the value of each parameter in a zone was essentially the average of 71 samples belonging to the concerned zone. Such a global average of each quality parameter is presented in Figure 2. The same figure also shows acceptable ranges of parametric values as stipulated by BIS (Table 1). The bar chart in Figure 3 shows the acceptable lower limit of a parameter as well as its upper limit of acceptance. It is apparent from Figure 3a that the regional average value of pH was higher than the acceptable lower limit of the pH value for potable water, whereas it lies below the upper allowable limit in respect to the specification laid down by BIS (Table 1). Similarly, average values of other parameters such as TA, TH, Cl, nitrate, fluoride, TDS, and residual chlorine were also mapped in the form of similar bar diagrams and are shown in Figure 3b–h. From Figure 3, it can be observed that the global average values for all the tested water quality parameters (eight in number) lie within their acceptable minimum and maximum values as per the BIS recommendation. When the value of a water quality parameter (as given by the average of 710 samples, divided into groups with 71 samples each) is found to be less than the permissible minimum or higher than the maximum permissible value as per specifications, it attracts attention from the users' side. It may be noted that the closeness of the average of a parameter towards the lower limit is different for different parameters; if we presume that the closer the value of a parameter is to the lower limit of acceptance, the better is the water quality in respect to the concerned parameter, it becomes apparent that individual quality parameters have different goodness of quality. This leads one to think of adopting a rational approach to describing the overall quality of water by linking the quality goodness of individual parameters. Incidentally, pH is somewhat an exception as it is universally accepted that a value of 7 represents the best pH value desirable in drinking water. However, there are other important water quality parameters, some of which are known to directly impact the pH value. So, to

rationalize the parametric contribution to determining the overall quality of drinking water, the lower limit acceptance of a specific quality parameter is considered its best possible quality goodness. In this respect, a pH value of 7.0 is taken to be the acceptable lower limit. It may be noted that there are other parameters that affect the water quality, for example, BOD, COD, dissolved oxygen, total coliform, and conductivity. In fact, there has been a report on water quality modeling using an ANN wherein as many as 56 input nodes were used [23]. With the aim to forecast algal growth in Tolo Harbour, Hongkong, Deng et al. [23] used an ANN of structure, $[56]_{\text{input}}-1-[1]_{\text{output}}$, and modeling was carried out in MATLAB. With four different algorithms, a learning rate of 0.01 and training epoch of 1000 was used to achieve better predictive power. Another machine learning (ML) technique, support vector machine (SVM), was also used to find out the suitability of a technique with respect to the forecasting capability [23]. RMSE and correlation coefficient R-values were used to judge the performance of various options. The performance of the SVM was better than that of the ANN, but with a higher computation time; of the different algorithms used in the ANN, the performance of the Levenberg–Marquardt (LM) algorithm was found to be superior. However, the present work deals with the formulation of the quality index of potable water. Though more input parameters, including those stated above, could have been used in the modeling, we felt it prudent to validate the conceived model with the use of these eight parameters, which appear to be quite important with respect to sensitivity towards determining the quality of drinking water. The use of more parameters would have given rise to a different result and could possibly be related to the drinkability of water. However, in that case, much more complexity would be involved in modeling as many of the parameters could have been found to be insignificant. Notwithstanding the experimental limitations, it would be an interesting exercise to model those parameters as well, and this work may be taken up as a separate study.

Accepting that different quality parameters have different goodness, it seems worth seeking a unique stochastic token that can describe the combined effect of all the water quality parameters and give the best idea for the quality of drinking water. In line with previous work, we propose defining such a stochastic token as a ‘water quality index’ (WQI) [24].

From the results in Table 2, the cumulative average values of each quality variable for the entire region are calculated by taking the average of the zonal average values of individual water quality parameters. Thus, the net average value of a quality parameter is equal to the sum of the average value of a parameter for each zone/total number of zones. Based on these derived average values, a simple algorithm is proposed to evaluate the individual parameter’s net quality index; taking the individual parameter’s quality index into consideration, the overall water quality index for the concerned region is described.

3.2. Water Quality Index

The present proposition of designing the water quality index for the Jodhpur region aims to qualify the degree of goodness of drinking water on a scale of 0–100. A value of 100 is obtainable for a parameter only if its average value equals the set lower limit, notwithstanding the achievable betterness below the set lower limit, which is called the permissible limit in the BIS standard. The measured data reveals that there is little scope to fix any other lower limit below the so-called permissible limit.

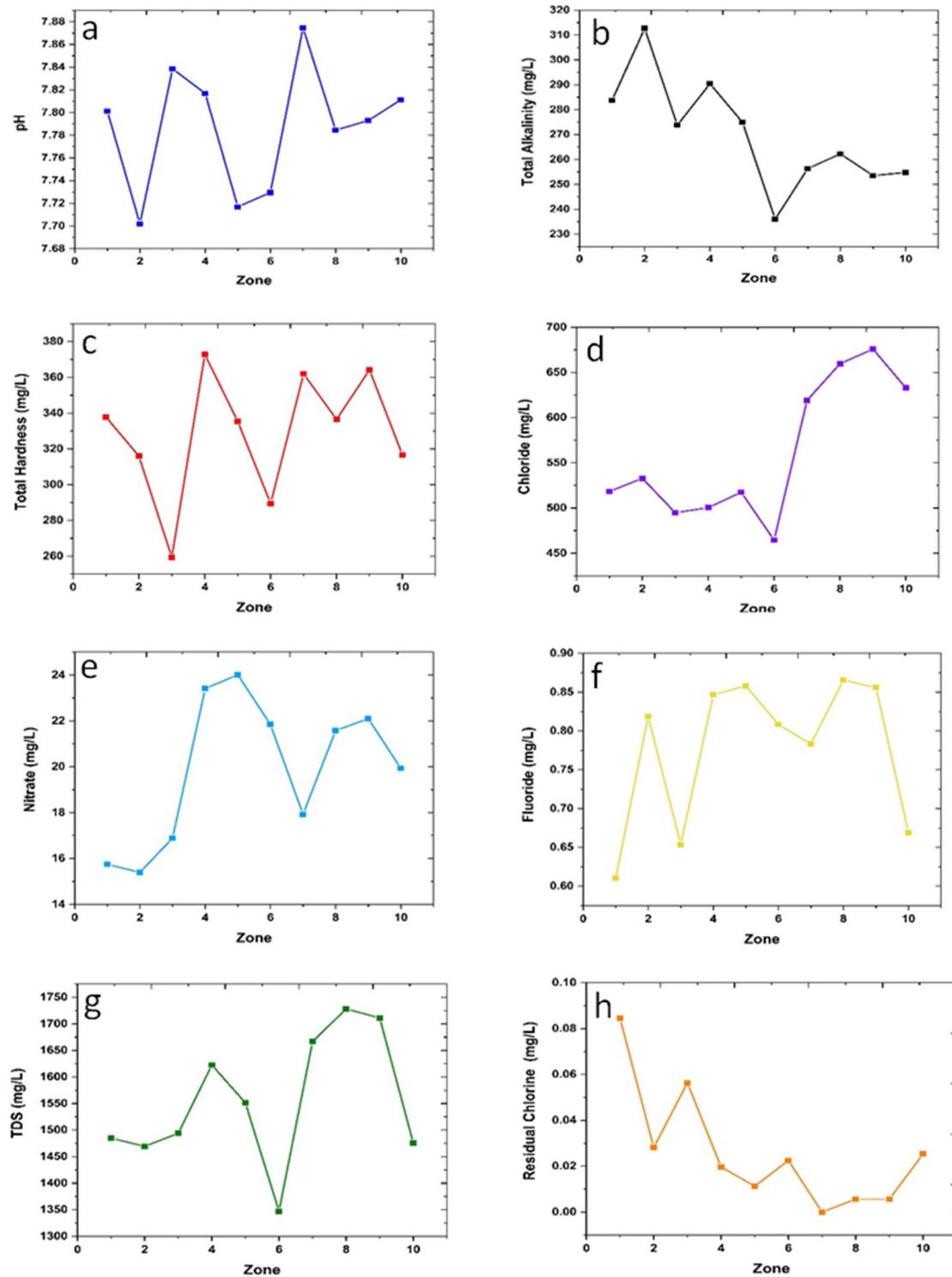


Figure 2. Variation of average water quality parameters with zones: (a) pH, (b) TA, (c) TH, (d) Cl, (e) nitrate, (f) fluoride, (g) TDS, and (h) residual chlorine.

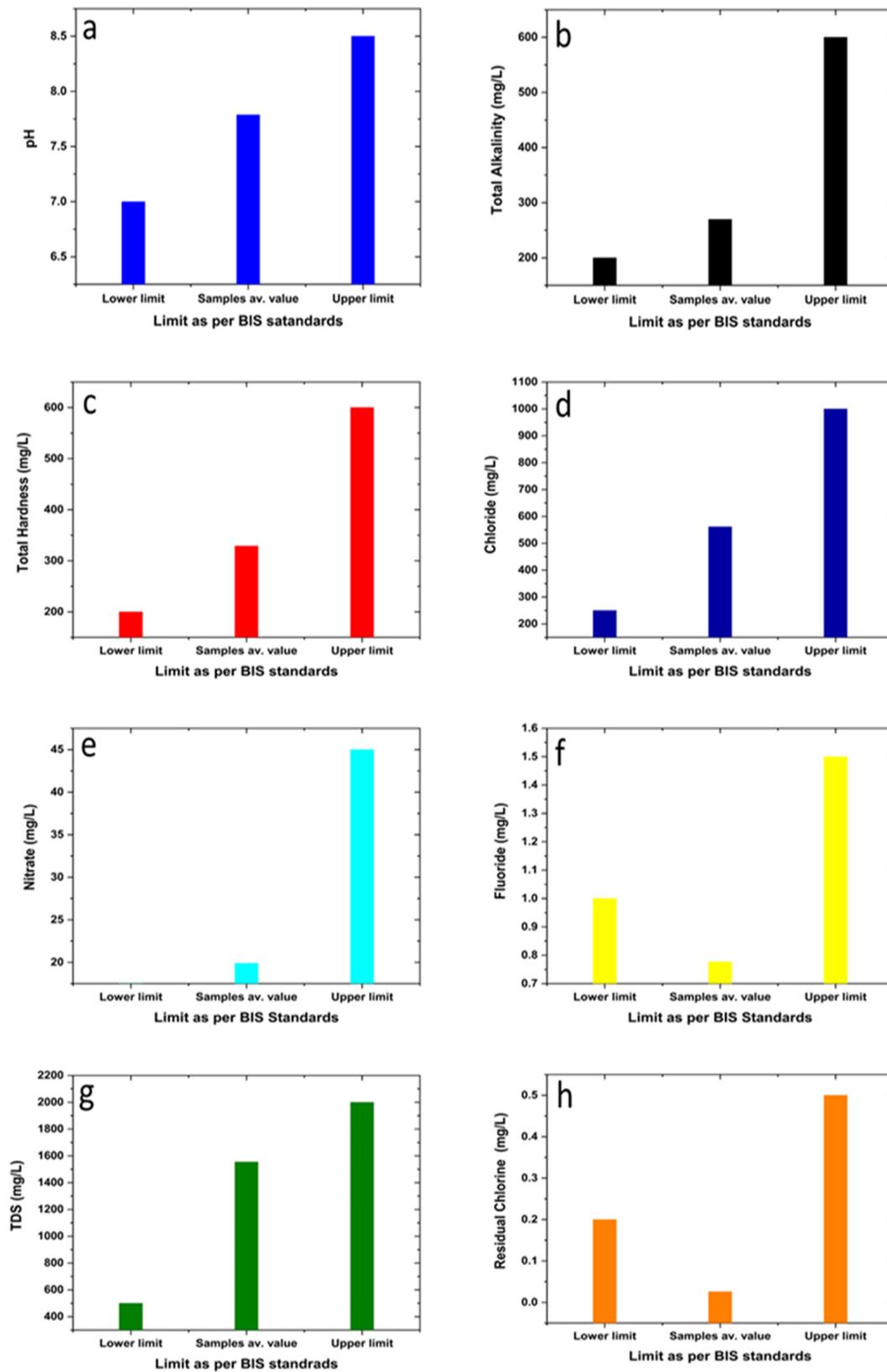


Figure 3. Average values of quality parameters over the entire experimental region against its acceptable limits as per BIS standards: (a) pH, (b) TA, (c) TH, (d) Cl, (e) nitrate, (f) fluoride, (g) TDS, and (h) residual chlorine.

The following algorithm is proposed for determining the water quality index of the chosen region in consideration of the results of 710 water samples tests, as carried out in the present research work (Figure 4). To accomplish this, the following assumptions are made:

- The acceptable lower and upper limit of quality parameters for use in the study are selected with an eye to the scope available in the BIS standard.
- Within a given range of specifications, the closer the average value (of a quality parameter) lies to the acceptable lower limit (Table 2), the higher the parametric quality index will be.
- If the minimum accepted value is not specified in a standard, the acceptable lower limit shall be considered zero.
- The proposition is generic and applies to any water quality standard that distinctly specifies the lower and upper limits of acceptability for a water quality parameter. The goodness or badness of parameter value beyond either limit is not considered as it does not fall within the scope of the study with water samples from the Jodhpur District of India.
- If the quality index comes out to be more than 1, it is to be taken as 1 (as a value lower than the minimum accepted value may present health concerns in some instances).
- The limits set by the model describe the best or worst goodness of water quality; if there lies any consequence, better or worse, beyond the prescribed quality limits, the same is not given weightage.

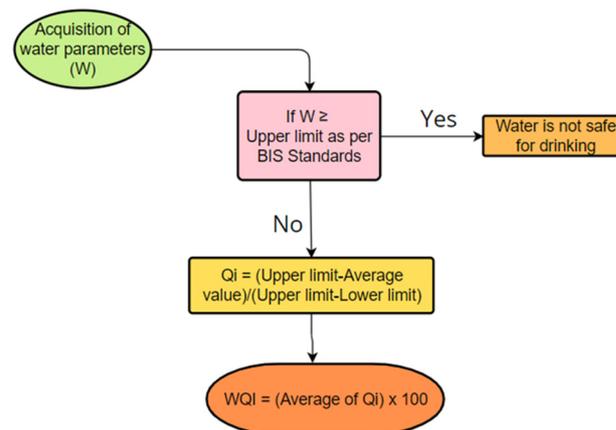


Figure 4. Flow chart to determine the water quality index.

It may be noted that the assumptions follow from the science of water based on which different specifications are laid down; while the use of other standards will give rise to the different absolute values of WQI, the proposed methodology to calculate the WQI will not be affected. In such cases, the quality gradation scale needs to be altered with respect to a scientifically branded ideal situation, such as 7 for pH.

Step 1: Call the average value of each water quality parameter as $a_1, a_2, a_3 \dots \dots a_n$.

Step 2: Denote the acceptable upper limit of each parameter as per BIS standards by $b_1, b_2, b_3 \dots \dots b_n$.

Step 3: Denote the minimum allowable limit (lower limit) of each parameter as per BIS standards by $c_1, c_2, c_3 \dots \dots c_n$.

Step 4: Compare each of the average values with its corresponding maximum limit.

Step 5: If any $a_i > b_i$, discard the unsafe water; else, go to the next step.

Step 6: If $c_i < a_i < b_i$, accept the water, and go to the next step and calculate the water quality index due to the i th parameter.

Step 7: For $c_i < a_i < b_i$, find the quality index of the i th parameter as

$$Q_i = \frac{\text{Upper limit} - \text{average value}}{\text{Upper limit} - \text{Lower limit}}$$

$$\text{hence, } Q_i = \frac{b_i - a_i}{b_i - c_i}$$

Step 8: Find the average value of all the individual quality indexes of each individual parameter and define it by water quality index (WQI) for the experimental region:

$$\text{WQI} = \sum_{i=1}^n \frac{Q_i}{n} \times 100$$

where Q_i stands for the quality index of an i th parameter over the entire region, and ' n ' is the number of the quality parameter in consideration (eight in the present work).

The WQI levels are also categorized as follows:

For WQI lying within:

90–100—The water is excellent for drinking.

70–90—The water is good for drinking.

50–70—The water is of medium quality but still safe for drinking.

25–50—The water is of a bad quality and unsafe for drinking.

0–25—The water is very bad and is highly unsafe for drinking.

Based on the above definition of the water quality index, the WQI has been calculated. The manner of defining a WQI by taking the quality goodness of individual parameters leads to an important research question to answer. It can be seen that the quality index (Q_i) of each parameter has its own weightage in the determination of the final WQI value. In this case, we have considered eight quality parameters and each parameter has 710 sampled measurement data which are used in the aforesaid manner to calculate WQI. Different quality parameters have a different impact on the ultimate water quality index. Hence, the final quality of water is determined by the combined effect of all eight quality parameters. It is very likely that these eight quality parameters (pH, alkalinity, TDS), taken together, bear a nonlinear relationship with the so-defined WQI. This means that the WQI is an unknown function of eight variables which are the above eight quality parameters. It may be noted that there are other standards that could be used for determining WQI. USEPA specifies the allowable limits very similarly to the BIS standard. However, it is important to note that the specification of parameters varies in different standards of different countries; WHO prescribes rationalized standards. Since the paper deals with the parametric limits set by a standard, the quality index value should change; in that case, the acceptability limit needs to be redesigned in harmony with the specification.

Given the variation in values of a quality parameter with respect to time and space, it seems interesting to model the relationship between the eight quality variables as input and the resultant WQI as the output. While there are many statistical tools to map the hidden relationship between the input variable and the resulting output, the artificial neural network is considered a powerful tool. In classical perceptron learning, a feed-forward backpropagation algorithm is used. The weighted inputs with a bias value are operated with a prechosen approximator (called the transfer function), and then the calculated output is compared with the target output. The alteration of the weight value minimizes the observed error in each iteration. With this understanding, we have implemented artificial neural network modeling to map the nonlinear relation between the water quality parameters as the input and the individual quality index as the output.

The selected neural network architecture is shown in Figure 5. It is a four-layer network comprised of an input layer with eight nodes (each node represents one quality parameter), two hidden layers, and the final output layer with the lone node representing the quality index (WQI). The first hidden layer consists of 10 nodes connected with every node of the second hidden layer, which contains 6 nodes. The connectivity between the second hidden layer and the output layer can be seen in Figure 5. It may be mentioned that one may use a different architecture of ANN; the number of hidden layers, number of neurons in a hidden layer, as well as the topology of the ANN structure can be varied. There are several network architecture protocols. However, an increase in the number of nodes

in a hidden layer and the number of hidden layers in a classical ANN does not necessarily guarantee good learning. Too much lowering of the mean square error value may lead to a situation when the ANN will learn only the pattern it is shown and will not be able to predict the outcome of a similar situation with a data set. That means the ANN may lack the capability of generalized learning. In this case, we tried to increase the number of the hidden layer from one to five with an increasing number of nodes; in the majority of the cases, the training error curve did not converge well and did not reach a low value reproducibly. Moreover, it was observed that the architecture $[8]_{(input)}-10-6-[1]_{(output)}$ could give us a relatively better result. The training, testing, and validation curves are found to be acceptable. Moreover, one could also adopt a different technique to more authentically optimize the ANN architecture by the use of the genetic algorithm. The required objective functions of the genetic algorithm may be obtained from a preceding multivariate analysis. However, this is not within the scope of the present research. The authors propose working on artificial learning of the interrelations of water quality parameters as a separate exercise.

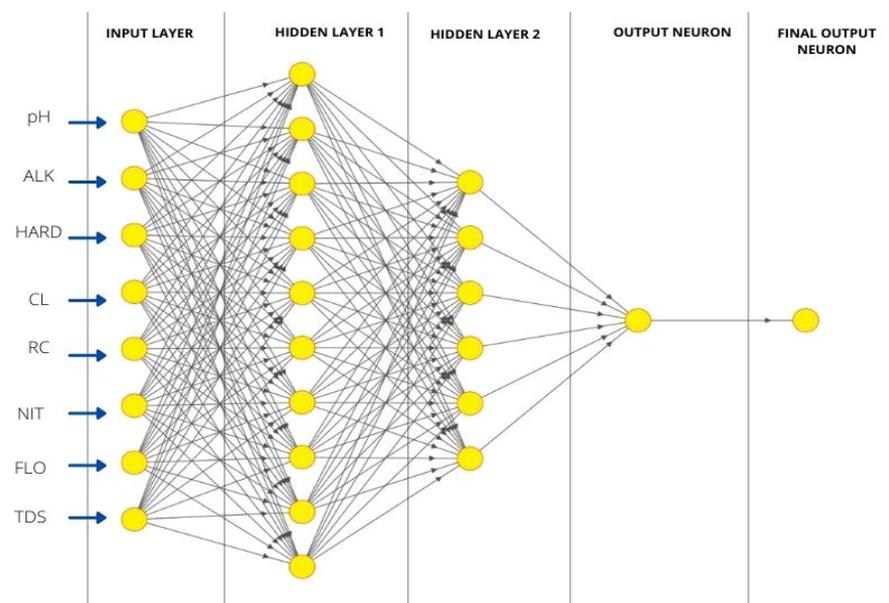


Figure 5. The artificial neural network architecture used for the present study.

Moreover, in any such case of relational pattern recognition, it is imperative to make use of an approximator, termed a transfer function in ANNs. When we work with the related toolbox in a MATLAB platform, we have to choose any of the given transfer functions available in the toolbox. One can, of course, invoke the use of a higher-order universal approximator as a transfer function, but it would require the writing of suitable codes both for the backpropagation algorithm and for getting it interfaced with the software it works with. However, this itself will be separate research without much surety of good convergence in the problem concerned. This activity is data sensitive and hence it is an educated game to rationalize the best possible approach. The authors do not claim that the adopted strategy is the best possible one for the present data set for water quality modeling. The task pursued here is the simplest way of ensuring the predictive power of an ANN. The MATLAB platform used the feed-forward backpropagation LM algorithm to train the network. As in a classical neural network, the general scheme of data flow is also followed in the present work and is shown in Figure 6. The data flow in the forward direction, which by the backpropagation algorithm changes the weight at each node, and the output is changed until training is stopped at a desirably low training error. It is apparent that, at each of the hidden layer nodes, the weighted input is added with a randomly chosen bias value before being put to the approximator used in the present case, which is the transfer function. The Tanh transfer function is known to be quite efficient in capturing

nonlinearity [25,26]. In contrast, the present ANN modeling does not use any existing knowledge about the effect of the individual parameter. It has undergone supervised learning with the intent of recognizing the hidden relational pattern among the quality index assignable to the individual parameter; this kind of exercise is entirely new to its kind as there is no example where the contribution of the quality indices of eight parameters is integrated through a well-known learning process. The LM algorithm used here has produced a relatively better correlation; the authors tried with other backpropagation algorithms available in the MATLAB toolbox. The gradient descent algorithm and scaled conjugate gradient (SCG) algorithm were also tried, but in vain. We have the provision of using only those algorithms which are available in MATLAB. There is no denying that different algorithms have a different propensity for learning curves being trapped in local optima. To secure a global optimum, one needs to adopt different techniques. As stated earlier, one such technique is to use a genetic algorithm with a preceding backup of multivariate analysis. The other approach to good learning could be the neuro-fuzzy techniques; one may also test the case with unsupervised learning through the Kohonen network. All said and done, individual activity is a large task by itself and the authors have chosen the simplest one to get the idea about how the individual quality indexes interplay with one another. Apart from this, one may also prefer to use a Bayesian neural network, autoregressive moving average, or decision support system; moreover, several other deep learning techniques could also be tested for better prediction. However, for the ANN used here, the performance is best judged by MSE, R-value, and R-square and is considered to be sufficient. For other processes such as K-nearest neighbor, KNN, SVM, or the naïve Bayes model, other parameters such as accuracy, sensitivity, specificity, and F-score are used to judge the performance. Moreover, other than the use of MATLAB for the neural network, other useable software include Tflern, Neural designer, Keras, Neuro Solution, Torch, and Microsoft Cognitive Toolkit. The neural designer can be used to mathematically model a similar data set in a code-free manner, enabling artificial intelligence (AI)-powered applications.

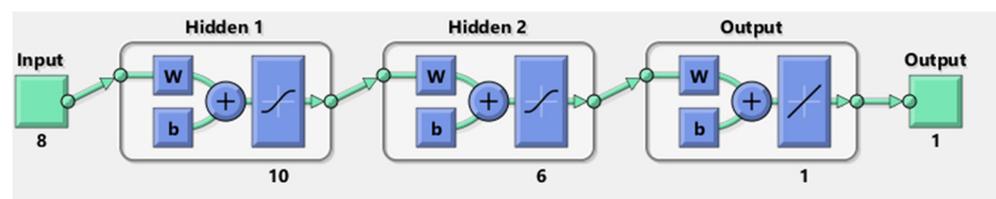


Figure 6. The general scheme for a feed-forward backpropagation perceptron learning.

All 710 of the measurement data for each input variable were considered for modeling. A total of 70% of the data was taken to train the neural network, whereas 15% of the data was used for testing, and another 15% was taken for validation of the model. Since the ANN modeling was carried out in the MATLAB platform, the selection of data for training, testing, and validation was automatically random. A code was written to further randomize the given data set to reinforce the observation from the ANN modeling in MATLAB. Such random data selection was performed a number of times, and for each randomly selected data set, training, testing, and the corresponding validation were performed.

Keras can be used for purposes such as convolutional neural networks (CNNs) and recurrent neural networks. Essentially, these are deep learning software that could be used for learning the problem of the present one. The authors contemplated using the deep learning software to have better introspection into the problem being handled. The present work is a preliminary investigation to explore the feasibility of using learning techniques simulating the human brain such that one can map the relational aspects among the various parameters. It is not out of context to refer to the elegant work of Kouadri et al. [27], where the performance of eight different machine learning techniques were used for predicting the water quality index; the artificial intelligence algorithms (AI) used by the authors were

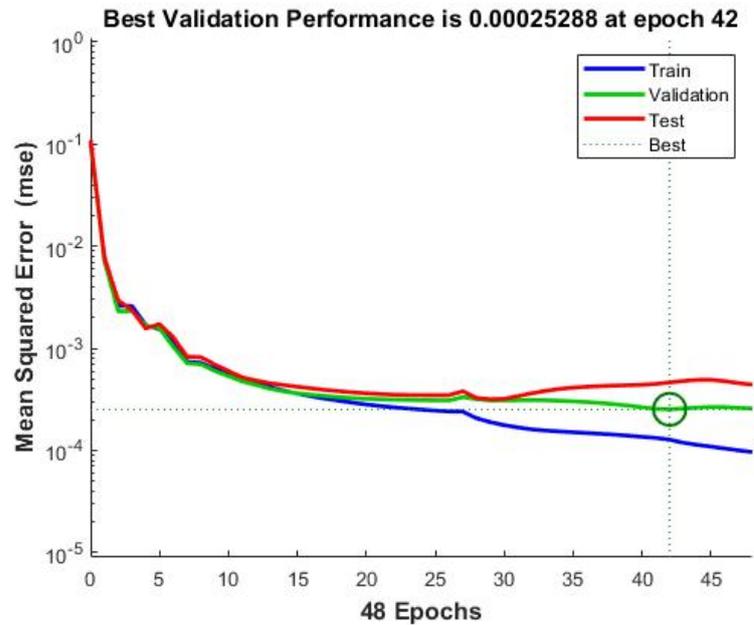
multilinear regression (MLR), support vector machine (SVM), artificial neural network (ANN), random forest (RF), random subspace (RSS) additive regression (AR), locally weighted linear regression (LWLR), and M5 P tree. While taking 12 inputs, the authors reported the superior algorithm. The authors used MATLAB for ANN and MLR, whereas for all other models, Waikato Environment for knowledge analysis (WEKA-version 3.8.4) was employed. The authors could find out the two most sensitive input parameters by sensitivity analysis, and these were further subjected to modeling, thereby observing the superiority of RF over the others; incidentally, ANN was found to be the second-best. When compared with the present work, it becomes evident that such a unique approach was used to evaluate the efficacy of ML techniques in understanding the water quality index of a particular parameter; the performance evaluation is the R-value, mean absolute error value (MAE), root mean square error value (RMSE), root-relative square error (RRSE), and relative absolute error. We contemplated a different task; after obtaining the values of quality indices of eight parameters that are considered to be a significantly important determinant of the potability of water, the unique water quality index (WQI) number for a specific geographic location is defined as per the proposed model of water quality indexing. We have been in search of a unique number that describes the water quality index in consideration of the individual parametric contribution to the overall water quality index. As has been advocated elsewhere [27], there is a difference in the relative sensitivity of a quality parameter. It is logical to assume that the overall WQI is a complex function of the individual's contribution (Q_i). To understand the hidden relationship, which presumes to be nonlinear, we resorted to the use of an ANN as a learning tool. Our approach is quite different from what is reported to date in respect to WQI modeling. Herein lies the novelty of our work.

Appreciating that there is a dependence of the WQI on the eight chosen water quality parameters, which assume different values for different samples drawn from different places or times, it appears to be an educated game for predicting the WQI for any set of water parameters. As a number of previous works have discussed, ANNs are one such powerful predictive tool (15–17, 21) in describing the water quality index amidst the changing water quality parameters. We have designed an ANN architecture and have performed training, testing, and validation repeatedly, each time with randomly selected data. This approach seems to be more practical than any sequential data selection strategy and is expected to avoid overoptimism. The representative performance plot of the neural network is shown in Figure 7a.

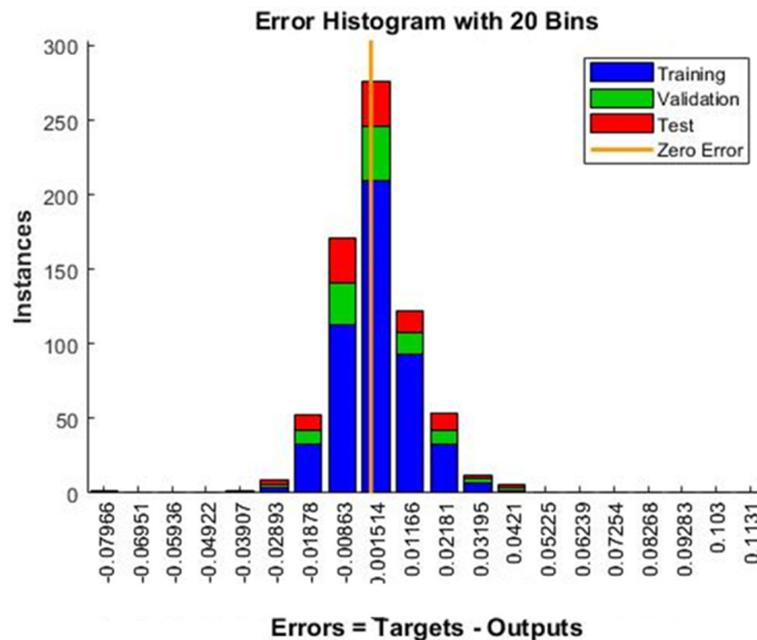
It can be observed that the training, testing, and validation error values gradually diminish, and the consistency in behavior can be noted in the figure. The validation performance is also quite good. It is evident from Figure 7a that the test error, validation error, and training error are rather close to one another, which signifies that the designed neural network used for learning the input–output relation ($WQI = f(\text{quality parameters})$) is rather reliable. It was also found that the best achievable validation performance was 0.00025288 and was obtainable at the epoch 42. Moreover, the error histogram of training, testing, and validation is shown in Figure 7b. It can be seen from Figure 7b that the error defined by the difference between the target and output values is distributed over a very narrow region; this observation is valid for training, testing, and validation. It is, therefore, apparent that the present neural network is capable of effectively learning the relation between input parameters and the final output. Hence, this ANN was then subjected to further performance assessment.

As stated in the preceding discussion, it is important to know the ability of the network to understand the relational behavior existing within the dataset, as well as the accuracy with which the ANN can predict the change in the WQI with changing values of water quality variables. The performance of the network is assessed by the correlation between the output and the target. The correlation curves obtained from the modeling in the MATLAB platform are shown in Figure 8. It may be noted that the correlation coefficient obtained from ANN modeling in the MATLAB platform always represents the Pearson

correlation coefficient. Figure 8 shows that the R-value was 0.98815 for training, 0.94917 for testing, 0.97243 for validation, and the overall correlation coefficient, the R-value, obtainable for all the data may be as high as 0.98071. The magnitudes of the R-value for training, testing, validation, and, finally, for all data, indicate a good generalization. From the observed results in Figure 8, it appears that the network’s performance is expectedly very satisfactory [28].



(a)



(b)

Figure 7. (a) Performance plot of the implemented ANN model. (b) Error histogram for training, testing, and validation.

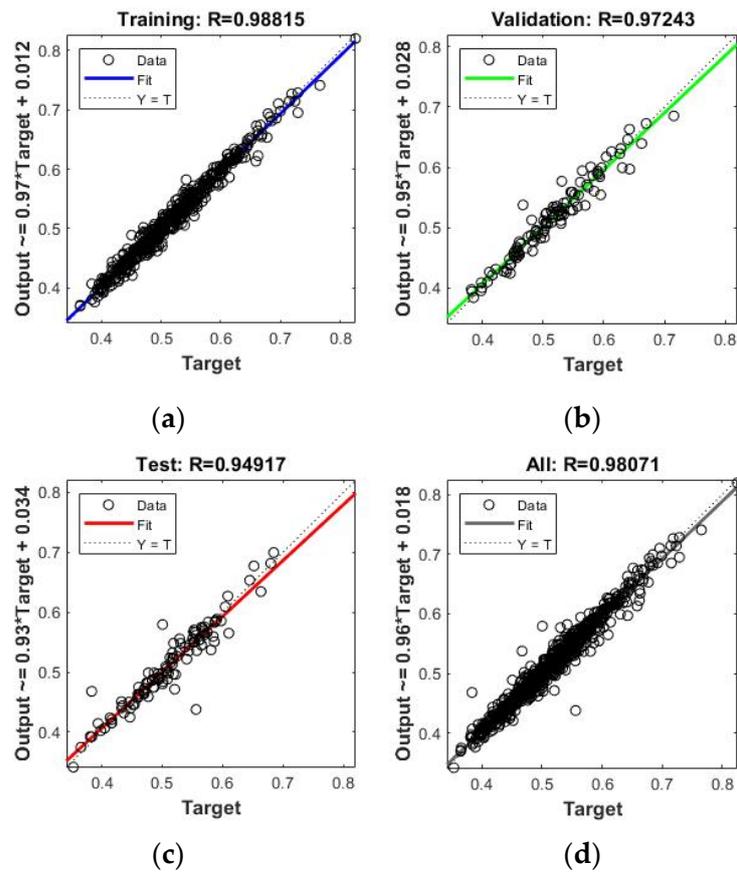


Figure 8. Correlation between the actual network output and targeted output: (a) training, (b) testing, (c) validation, and (d) final output.

It is known that the learning activity in a neural network involves the training of the network, during which the mean square error (MSE) is normally seen to decrease with increasing iterations. Overfitting of data may result in poor generalization; this means the network will recognize the pattern shown to it, and it will not be able to get into generalized learning. Too low an MSE value is not always desirable as it signifies that the network is trained to recognize the pattern shown to it. However, the objective is to acquire generalized learning. Hence, the network training is stopped when the desirable MSE is obtained and the performance of the network is subjected to assessment. While testing the model performance, the average value of the R2 and root mean square error (RMSE) is usually examined. As stated above, the overall model performance is quite good as one can see that the model output and the target value of the WQI bear a correlation coefficient of 0.98071, implying that the neural network has satisfactorily learned the interrelation between the input variables and the output.

While the nearness of the R-value for training, testing, and validation indicates a good generalization, it is also essential to introspect into the training state situation to assess the overall output capability of the ANN. The training state plot for the designed network is shown in Figure 9.

It can be noticed from the gradient coefficient that the learning rate for the learning process adopted is rather low; moreover, the validation fails against the number of epochs showing that validation checks equal 6 after 48 epochs. However, all these performance indicators of the designed ANN are presented in summarized form in Table 3.

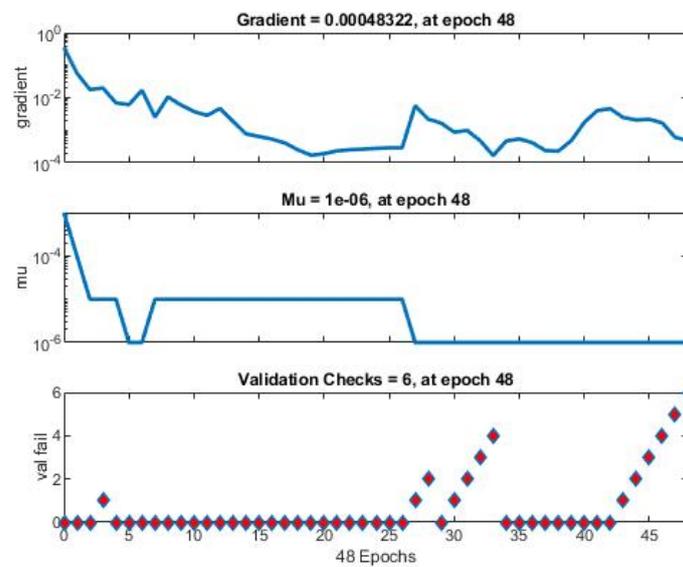


Figure 9. Training state plot of the designed ANN until epoch 48.

Table 3. The output of the experimental neural network.

Parameters	ANN [8]-10-6-[1]
Gradient	0.00048 at epoch 48
Validation fail	6 at epoch 48
Learning rate	1×10^{-6}
Training R	0.988
Validation R	0.972
Test R	0.949
All R	0.980

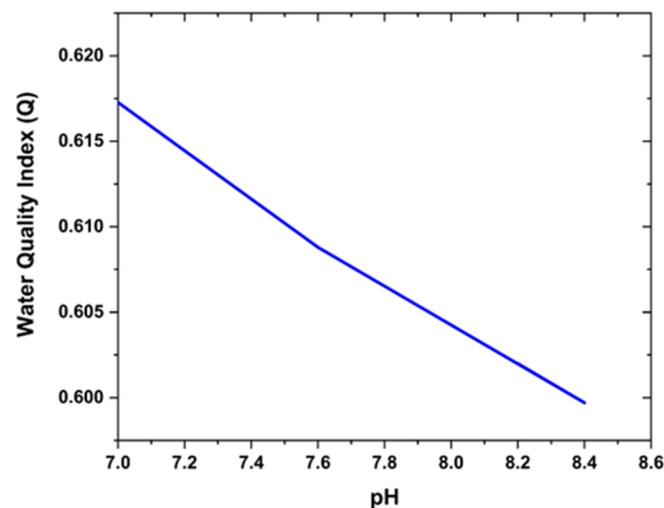
It can be observed from Table 3 that the gradient coefficient of 0.00048, a learning rate of 1×10^{-6} , is achievable at a correlation coefficient value of 0.980. As stated earlier, the entire task is repeated several times with a random selection of data set at each time so that the average value can reasonably say that the model is stable. The results of ten such meaningful results are presented in Table 4.

Table 4. Performance of the neural network used for WQI prediction.

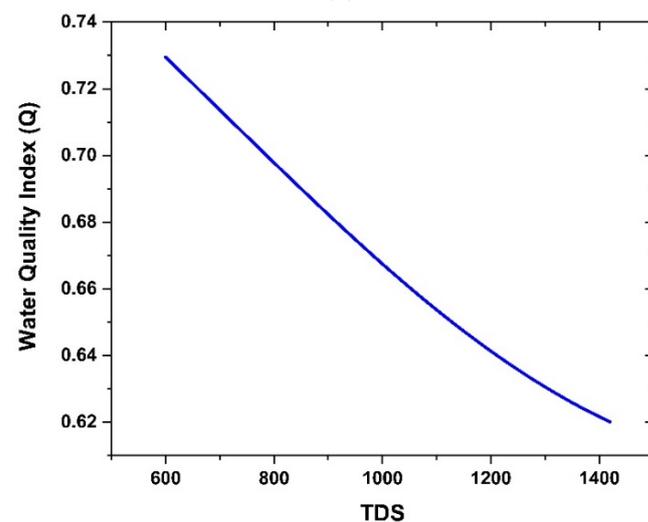
Reading	MAE	RMSE	RSQR	TRAIN PERF	VAL PERF	TEST PERF
1	0.000458	2.10×10^{-7}	0.9618	0.000144	0.000458	0.000548
2	0.000516	2.66×10^{-7}	0.9598	0.000127	0.000516	0.00028
3	0.000344	1.18×10^{-7}	0.9632	0.000114	0.000344	0.001008
4	0.000478	2.28×10^{-7}	0.9589	9.66×10^{-5}	0.000478	0.000407
5	0.000426	1.82×10^{-7}	0.9624	9.14×10^{-5}	0.000426	0.00026
6	0.000345	1.19×10^{-7}	0.9629	0.00019	0.000345	0.000879
7	0.000296	8.77×10^{-8}	0.9599	0.000226	0.000296	0.000262
8	0.000328	1.08×10^{-7}	0.9623	0.000159	0.000328	0.000608
9	0.000414	1.71×10^{-7}	0.9614	0.000182	0.000414	0.000429
10	0.000144	2.07×10^{-8}	0.9633	5.25×10^{-5}	0.000144	0.000183
Average	0.000375	1.51×10^{-7}	0.96159	0.000138	0.000375	0.000486

It can be observed that the performance parameters are quite consistent for all the ten datasets; this contributes to authenticating that the model is stable. Moreover, the performance indices of the present ANN model are compatible with similar models reported elsewhere [17–20]. It may be noted that the average value of the coefficient of multiple determination, viz., R^2 was 0.96159; the standard deviation of R^2 was also calculated and

was found to be 0.00155. This authenticates the stability of the model. Generalized learning by an ANN is of extreme importance; for this reason, the predicted dependence of WQI on the input variables needs to be mapped. This is done by varying a single quality parameter at constant values of all other parameters and then finding out the WQI as the network output. Therefore, the ability for generalized learning of the ANN was verified by the results of network prediction as revealed in Figure 10a,b. Figure 10a shows that the network predicted the plot of variation in quality index with the pH value. It was found that increasing pH value leads to deterioration of the water quality index. This is in agreement with the existing knowledge on the effect of the pH value of water on its potability. Likewise, Figure 10b shows how the water quality index decreases with increasing TDS. There is no denying that TDS affects the drinkability of water, and the results obtained from the use of the neural network are compatible with a well-established understanding in this regard. Therefore, it was found that the predictive capacity of the artificial neural network may be harnessed to monitor the water quality at any instant, and this finally helps in taking remedial steps to restore water quality through proper treatment.



(a)



(b)

Figure 10. (a) Variation of water quality index vs pH. (b) Variation of water quality index against TDS content in water.

4. Conclusions

It is concluded that the quality of water, as determined by various quality parameters, can be measured by a singular parameter that combines the effect of all parameters adversely affected by the inevitable contamination of source water by industrial wastewater. The present work results corroborate the observation of the previous works researching the ANN modeling of water quality. The authors further infer that a suitably designed unique stochastic token called herein a water quality index could be successfully used to measure the overall water quality, thereby enabling the knowledge of the role of industrial wastewater in polluting source water. The authors further conclude that the water quality index of the Jodhpur region in India is rather good (~60% or above) as per the newly introduced water quality indexing protocol, and it verifies that potable water in the experimental region is safe for drinking. The artificial neural network can predict the effect of individual quality parameters on the overall quality index of water amidst continuous contamination by industrial wastewater. The network prediction of the input-output relation is satisfactory. The authors conclude that prediction from neural network modeling can be employed to control the water quality by suitable means such as solar energy. However, the present work is rather concise and has left out several other options for better authentication of its outcome. As some more algorithms could be tested in the same platform with different architectures to reason out the availability of any better alternative in ANN modeling in the MATLAB platform, the adoption of other learning techniques including the new generation statistical modeling deserves specific attention in water quality modeling. It is encouraging to note that the results obtained in this study are comparable with previous observations and that there is a visible qualitative match as far as ANN modeling is concerned. Taking a clue from this exploratory, investigative work, the authors consider it prudent to extend this work by using other predictive techniques in machine learning. Unsupervised learning through data clustering may not be underrated. The ANN itself will be made an expert by incorporating the knowledge of physical sciences underlying the influence of individual parameters and mutual interactions among them on the potability of water. Synthesizing the knowledge created by us with those gathered from the literature, one may conclude that the present work has excellently opened up a newer horizon in the research on water quality index modeling by emerging machine learning techniques. In light of the above observations, the authors wish to conclude that the presently employed ANN technique can act as an effective tool to understand the hidden nonlinear relationship between the concentrations of water quality parameters and a well-defined water quality index within a specific geographic location.

Author Contributions: Conceptualization, K.K.S., M.K.G. and M.K.B.; methodology, K.K.S., M.K.G., M.K.B., S.K.S., G.M. and M.F.; software, K.K.S., M.K.G. and M.K.B.; validation, K.K.S., M.K.G. and M.K.B.; formal analysis, K.K.S., M.K.G., M.K.B., S.K.S., G.M., M.F., S.K., P.K. and N.S.; investigation, K.K.S., M.K.G., M.K.B., S.K.S., G.M., M.F., S.K., P.K. and N.S.; resources, M.K.B. and P.K.; data curation, K.K.S., M.K.G. and M.K.B.; writing—original draft preparation, K.K.S., M.K.G. and M.K.B.; writing—review and editing, K.K.S., M.K.G., M.K.B. and G.M.; formal analysis, K.K.S., M.K.G., M.K.B., S.K.S., G.M., M.F., S.K., P.K. and N.S.; visualization, K.K.S., M.K.G., M.K.B., S.K.S., G.M. and M.F.; supervision, M.K.G. and S.K.S.; project administration, M.K.G. and S.K.; funding acquisition, P.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data shall be provided by the corresponding authors upon special request.

Acknowledgments: The authors are thankful to the three anonymous reviewers whose critical review has improved the quality of this manuscript. We are also thankful to the editor for their insightful comments that helped to provide the final shape to this manuscript. The author G.M. is thankful to the Department of Science and Technology, Government of India (DST-Gol) for providing the Fellowship under the Scheme for Young Scientists and Technology (SYST-SEED) [Grant no. SP/YO/2019/1362(G) & (C)].

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Meraj, G.; Kanga, S.; Kranjčić, N.; Đurin, B.; Singh, S.K. Role of Natural Capital Economics for Sustainable Management of Earth Resources. *Earth* **2021**, *2*, 622–634. [CrossRef]
2. Shyam, M.; Meraj, G.; Kanga, S.; Sudhanshu; Farooq, M.; Singh, S.K.; Sahu, N.; Kumar, P. Assessing the Groundwater Reserves of the Udaipur District, Aravalli Range, India, Using Geospatial Techniques. *Water* **2022**, *14*, 648. [CrossRef]
3. Bhat, S.A.; Meraj, G.; Yaseen, S.; Pandit, A.K. Statistical Assessment of Water Quality Parameters for Pollution Source Identification in Sukhnag Stream: An Inflow Stream of Lake Wular (Ramsar Site), Kashmir Himalaya. *J. Ecosyst.* **2014**, *2014*, 898054. [CrossRef]
4. Himesh, B.Y.; Rao, C.V.C.; Mahajan, A.U. Application of Steady State Stream Water Quality Model. *J. IAEM* **2000**, *27*, 33–37.
5. Bi, S.; Wang, L.; Li, Y.; Zhang, Z.; Wang, Z.; Ding, X.; Zhou, J. A Comprehensive Method for Water Environment Assessment considering Trends of Water Quality. *Adv. Civ. Eng.* **2021**, *2021*, 5548113. [CrossRef]
6. Han, Z.; Zhang, Y.; An, W.; Lu, J.; Hu, J.; Yang, M. Antibiotic resistomes in drinking water sources across a large geographical scale: Multiple drivers and co-occurrence with opportunistic bacterial pathogens. *Water Res.* **2020**, *183*, 116088. [CrossRef]
7. Bhat, S.A.; Meraj, G.; Yaseen, S.; Bhat, A.R.; Pandit, A.K. Assessing the impact of anthropogenic activities on spatio-temporal variation of water quality in Anchar lake, Kashmir Himalayas. *Int. J. Environ. Sci.* **2013**, *3*, 1625.
8. Cui, M.; Sun, Y.; Huang, C.; Li, M. Water Turbidity Retrieval Based on UAV Hyperspectral Remote Sensing. *Water* **2022**, *14*, 128. [CrossRef]
9. Meraj, G.; Romshoo, S.A.; Ayoub, S.; Altaf, S. Geoinformatics based approach for estimating the sediment yield of the mountainous watersheds in Kashmir Himalaya, India. *Geocarto Int.* **2017**, *33*, 1114–1138. [CrossRef]
10. Puri, P.J.; Yenkie, M.K.N.; Sangal, S.P.; Gandhare, N.V.; Sarote, G.B.; Dhanorkar, D.B. Surface water (Lakes) quality assessment in Nagpur city (India) based on Water quality index (WQI). *Rasayan J. Chem.* **2011**, *4*, 43–48.
11. Bhat, S.A.; Meraj, G.; Pandit, A.K. Assessing the influence of stream flow and precipitation regimes on water quality of the major inflow stream of Wular Lake in Kashmir Himalaya. *Arab. J. Geosci.* **2015**, *9*, 50. [CrossRef]
12. Heidtke, T.M.; Sonzogni, W.C. Water Quality Management for the Great Lakes. *J. Water Resour. Plann. Manag.* **1986**, *112*, 48–63. [CrossRef]
13. Chandra, S.; Singh, A.; Tomar, P.K. Assessment of Water Quality Values in Porur Lake Chennai, Hussain Sagar Hyderabad and Vihar Lake Mumbai, India. *Chem. Sci. Trans.* **2012**, *1*, 508–515. [CrossRef]
14. Lung, W.S. Lake Acidification Model: Practical tool. *J. Environ. Eng.* **1987**, *113*, 900–915. [CrossRef]
15. Graf, R.; Zhu, S.; Sivakumar, B. Forecasting river water temperature time series using a wavelet–neural network hybrid modelling approach. *J. Hydrol.* **2019**, *578*, 124115. [CrossRef]
16. Setshedi, K.J.; Mutingwende, N.; Ngqwala, N.P. The Use of Artificial Neural Networks to Predict the Physicochemical Characteristics of Water Quality in Three District Municipalities, Eastern Cape Province, South Africa. *Int. J. Environ. Res. Public Health* **2021**, *18*, 5248. [CrossRef]
17. Ubah, J.I.; Orakwe, L.C.; Ogbu, K.N.; Awu, J.I.; Ahaneku, I.E.; Chukwuma, E.C. Forecasting water quality parameters using artificial neural network for irrigation purposes. *Sci. Rep.* **2021**, *11*, 24438. [CrossRef]
18. Du, Z.; Qi, J.; Wu, S.; Zhang, F.; Liu, R. A Spatially Weighted Neural Network Based Water Quality Assessment Method for Large-Scale Coastal Areas. *Environ. Sci. Technol.* **2021**, *55*, 2553–2563. [CrossRef]
19. Gupta, R.; Singh, A.N.; Singhal, A. Application of ANN for Water Quality Index. *Int. J. Mach. Learn. Comput.* **2019**, *9*, 688–693. [CrossRef]
20. Aldhyani, T.H.H.; Al-Yaari, M.; Alkahtani, H.; Maashi, M. Water Quality Prediction Using Artificial Intelligence Algorithms. *Appl. Bionics Biomech.* **2020**, *2020*, 6659314. [CrossRef]
21. Sulaiman, K.; Ismail, L.H.; Razi, M.A.M.; Adnan, M.S.; Ghazali, R. Water Quality Classification Using an Artificial Neural Network (ANN). *IOP Conf. Ser. Mater. Sci. Eng.* **2019**, *601*, 012005. [CrossRef]
22. Drinking Water Quality Standards (Drinking Water IS 10500: 2012). Available online: <https://iitk.ac.in/iwd/wq/drinkingwater.htm> (accessed on 31 December 2021).
23. Deng, T.; Chau, K.W.; Duan, H.F. Machine learning-based marine water quality prediction for coastal hydro-environment management. *J. Environ. Manag.* **2021**, *284*, 112051. [CrossRef] [PubMed]
24. Tyagi, S.; Sharma, B.; Singh, P.; Dobhal, R. Water Quality Assessment in Terms of Water Quality Index. *Am. J. Water Resour.* **2013**, *1*, 34–38. [CrossRef]
25. Datta, S.; Banerjee, M. Mapping the input–output relationship in HSLA steels through expert neural network. *Mater. Sci. Eng. A* **2006**, *420*, 254–264. [CrossRef]

26. Datta, S.; Banerjee, M.K. Kohonen network modeling of thermomechanically Processed HSLA steels. *ISIJ Int.* **2004**, *44*, 846–851. [[CrossRef](#)]
27. Kouadri, S.; Elbeltagi, A.; Islam, A.R.M.T.; Kateb, S. Performance of machine learning methods in predicting water quality index based on irregular data set: Application on Illizi region (Algerian southeast). *Appl. Water Sci.* **2021**, *11*, 190. [[CrossRef](#)]
28. Hadjisolomou, E.; Stefanidis, K.; Herodotou, H.; Michaelides, M.; Papatheodorou, G.; Papastergiadou, E. Modelling Freshwater Eutrophication with Limited Limnological Data Using Artificial Neural Networks. *Water* **2021**, *13*, 1590. [[CrossRef](#)]