International Journal of Computer Science and Engineering (IJCSE)
ISSN (P): 2278–9960; ISSN (E): 2278–9979
Vol. 15, Issue 1, Jan–Jun 2026; 13–26

International Academy of Science, Engineering and Technology
IASET
Connecting Researchers; Nurturing Innovations

AI DRIVEN WATER MANAGEMENT STRATEGIES FOR JODHPUR. A MACHINE LEARNING APPROACH

Dr. Ranjeeta Mathur[1] & Dr. Prabhat Mathur[2]

[1]*Associate Professor, Department of Life Science (Zoology), Lachoo Memorial College of Science and Technology, Autonomous, Jodhpur*

[2]*Associate Professor, Department of Computer Science, Aishwarya College of Education, Jodhpur*

ABSTRACT

In this paper, Using 710 samples taken in the water-scarce Indian city of Jodhpur in the state of Rajasthan, this article determines the water potability of a water supply by calculating eight water quality (WQ) metrics. We split the whole sample into 10 groups, each representing a distinct region. In order to describe the water quality for potability use, eight WQ parameters were chosen according to the methodology specified by the American Public Health Association (APHA). Depending on the zone, we look at how well each parameter performs. The whole water quality was described by a single number that was calculated by averaging the parametric values of several zones. It was discovered that there is a substantial variation in the average value of each parameter between zones. Our next step was to simulate the nonlinear connection between the water quality index and the aforementioned eight parametric inputs using machine learning techniques. There is evidence that the NN developed in this study learned enough to make accurate predictions about the input-output relationship. Additionally, it is clear that the WQI developed from this study is an effective tool for evaluating water quality in the research region. While the precise number is up for disagreement, this research takes a fresh approach to the problem of comprehending the cumulative influence of the many aspects influencing water quality, which is the biggest obstacle to providing a distinctive description of water quality for human consumption. Government agencies will benefit from this work's framework since it can be automated with the right technology and it will help them understand how water quality is changing so they can better control it.

KEYWORDS: *Water Quality Parameters; BIS Standards; Water Quality Index; Neural Network, Machine Learning Approaches, Nature Inspired Algorithm's.*

Article History

Received:* *04 Apr 2026* | *Revised:* *06 Apr 2026* | *Accepted:* *09 Mar 2026

INTRODUCTION

Animals and plants rely on water for survival, and it is also the most consumed substance by humans. River and lake water quality is crucial because it affects human health and the ability of aquatic species to survive. The pollution of source water by industrial effluent is a major concern for water quality, especially in metropolitan areas [1]. In addition, the main factor that determines ecological balance in the living world is the water quality in these bodies of water [2]. On the other hand, the quality of water has been declining due to the ongoing improvement in its use from natural sources [3].

It is well-known that the concentration of distinct quality parameters determines the characteristic quality of water needed for drinking, industrial operations, or agricultural. Consequently, different human activities need different parametric values of quality variables. As a result of persistent pollution from haracterizing on and the careless use of chemical haracteriz and pesticides in farming, aquatic biota has been steadily declining. In addition, people are contracting water-borne ailments due to the use of polluted water [4].

The suitability of water for a certain activity is determined by a number of technical characteristics. Nowadays, people are quite worried about the purity of the water they consume. There will be a worldwide water crisis unless new technologies are used to address the growing human population, the rising demand for water for varied reasons, and the inevitable wastewater creation [5].

The majority of nations have come up with creative solutions to reduce water pollution in their natural resources and have developed new methods of purifying water. Because of the international nature of many resource flows, a collaborative and inclusive mitigation plan is required. Because of this haracteriz, protecting Earth's water supply is now a top priority in science and technology [6]. The quality of water for human consumption, aquatic life support, and many agricultural and industrial uses is defined by its physical, chemical, and biological properties [7].

The appropriateness of water usage can be determined by ensuring that certain parameters, as listed in multiple studies, fall within specific ranges. In a water sample, certain metrics are probably going to be great and others won't be. Several scholars have argued that water should have a distinct identity to reflect its overall quality, in light of this variance. The water quality index (WQI) is described using various numerical values in this way [10,11]. Whatever its precise meaning, the water quality index has always served to clarify whether or not a particular body of water is appropriate for a certain use.

Water quality planning and management has been the subject of documented study, and a two-layer time-variable model has been created to measure seasonal changes in pH and alkalinity levels [13] for individual situations.

The estimation of pollutant loadings in water has made use of mathematical haracte, computer simulations, and other statistical haracte tools to forecast the deterioration of water quality across time and geography. The water quality index serves as the basis for its definition, and the studies use a decision support system to determine the likely solution's success. The evaluation of potable water quality via appropriate indexing is among the many essential uses of water quality indices.

Potable water is defined by the WHO and the BIS, who have established desirable levels for a number of water quality characteristics. There has to be proper treatment of the water sources that are contaminated by industrial wastewater discharge before they can be used again. This calls for a fresh approach to calculating the water quality index, yet one that is both easy to understand and use. It is feasible to establish the acceptable upper and lower limits of water quality parameters that are in harmony with the available parametric range within the research region, which is the Jodhpur district of India, using criteria given by several bodies. There is no intention of violating specifications or complying with rules in this research, which aims to build a water quality index in a more pragmatic but yet clear manner.

The WQI is a unified metric that several researchers have used to measure water quality. The input-output connection is inherently nonlinear, but current methods do not show it. As a result, we try to solve this problem by tracing the nonlinearity in the connection between the parametric quality contribution and the overall quality index. There is a lack of extensive documentation in the literature about the challenges of capturing nonlinear relationships between input and output sets and haracteriz related patterns using computational approaches. Modelling water quality as a function of several factors has attracted a lot of interest from academics because of the relevance of monitoring and forecasting the changing water quality [15].

A radial basis function (RBF) based on a neural network and multilayer perceptron learning were both used in an article on ANN haracte [16]. Effective management of South Africa's water resources has allegedly been made possible by a predictive capability. And in other places, ANN has been used to forecast water quality parameters for a year, thus improved management over irrigation water quality is possible. Results showed that the ANN models haracte for this task were able to accurately forecast the water quality [17].

When it comes to spatial and temporal assessment, the nonstationary nature of coastal water poses a significant challenge. Researchers have reportedly successfully. A more accurate distribution of water quality over the whole research area may be predicted using this method [18]. An ANN was also used to generate a water quality index. The experiment may have employed water from the Indian subcontinent, but the findings are applicable worldwide [19].

It was shown that the regression value might be improved at a higher learning rate with a larger training dataset. According to [19], the most effective model was a five-layer network. There is a wealth of well-documented literature on the topic of using AI algorithms to forecast water quality index [20]. The article detailed the creation of an algorithm for deep learning with short-term memory and an artificial neural network model. In addition, 3 machine learning algorithms—support vector machine, naïve Bayes, and K-nearest haracte (KNN)—were cited by the same group. All was OK with the models. An early effort was made to test if ANN could categorise water quality measures, given its shown pattern recognition capacity. We trained and tested using data from measurements of pH and dissolved oxygen, two water quality measures; the result was an 80% success rate in identifying data from these parameters with an RMSE of 0.468 [21]. From what has been said thus far, it is clear that there are a lot of additional learning approaches and approximators that may be utilized however, in order to do so, one has to build new code and ensure that it can interface with MATLAB without any problems. So, as a first step, it seems sense to employ neural network haracte to plot the relationship between commonly used drinking water quality metrics and a clearly defined water quality index. As shown before, this is consistent with the assertions made by the preceding writers. In order to make sense of the data at hand and determine how different parameter values affect water potability, the authors decided to use an artificial neural network (ANN) method to water quality haracte.

Consequently, this study employs neural network haracte using data on measured drinking water from a particular area. The water quality characteristics obtained at several sites in Jodhpur, an area of India that is water-stressed, were selected for neural network haracte in this research.

A natural resource that is fundamental to human haracterizi in every way, water is used for drinking, household purposes, agricultural purposes, and industrial processes. To make wise and efficient use of this valuable resource, water quality and quantity are critical. Climate change, or global warming, has been a result of human-caused processes in the last few decades (IPCC 2021).

A more often and severe occurrence of extreme weather events is one way in which climate change impacts the natural world. There are ambiguities around these consequences, and they have worsened the global water problem.

A SERIES OF LINEAR REGRESSIONS

Minimising the total squared error is the overall goal of multiple linear regression (MLR), a statistical haracte technique that seeks to create a linear connection between the dependent variable and several independent variables (Schneider et al. 2010). The MLR is straightforward and easy to comprehend, but it relies on a number of assumptions, such as that the connections between the dependent and independent variables are linear, that the observations are independent, that the residuals are normal, and that there is no multi-collinearity. Regardless of these drawbacks, MLR is still widely used in model construction for ML model comparisons.

RANDOM FOREST MODEL

Random Forest (RF) is a decision tree representation of a bagging technique, where. All of the various trees' outputs are weighted and then averaged to get the final forecast (Bilali et al. 2021). The decision tree and other models have been outperformed by this ML tool, which was first suggested by Breiman (2001), in several instances. Several WQI research also found that RF models performed better. When building the RF model, you need to define two parameters: max_depth, the number of trees to be included in the forest, and n_estimators, the number of variables in the random subsets at each node (Liaw & Weiner 2002). According to Shams et al. (2024), grid search is a useful tool for haracteri the hyperparameters of ML models used for water quality prediction.

FINDINGS AND ANALYSIS

The average values of several water quality measures varied across different sites, as shown in Figure 2a-h of the water quality evaluation. The zones were chosen at random and do not indicate any clear functional relationship with distance or other demographic or geographic factors. Therefore, the quality characteristics differ depending on the location, as illustrated in The average values of the quality parameters show a large amount of diversity across the chosen zones. Figure 2a shows that the pH value varies among zones; it's higher than the ideal value of 7, which is 7, and falls anywhere between 7.70 and 7.88. A other criterion, total alkalinity (TA), also fluctuated from similarly, with 0.2 established as the minimum allowable chlorine level. It can be harmful to your health if its value drops below this point. The same logic applied to fluoride; regardless of whether lower levels are beneficial to health, the lower limit was set at 1.

Considering that the average of seventy-one samples per zone changed depending on where the zones were located, it makes sense to test whether the sum of the zonal averages of each parameter may be useful in estimating the region's water potability as a whole. Taking the average of the individual quality averages for each zone allowed us to get the overall average value of each water quality parameter for the whole area. Each parameter in a zone was effectively calculated by averaging the results of 71 samples from that zone. What you see in Figure 2 is the worldwide mean of all the quality metrics. Table 1 of the same figure provides the permissible parametric value ranges as specified by BIS. Figure 3's bar chart displays the allowed range for a parameter, including both its minimum and maximum values. Figure 3a shows that the regional average pH was more than the minimum allowed value for potable water, but it is lower than the maximum permitted value according to BIS specifications.

The worldwide average values for all eight of the water quality metrics that were evaluated fall within the acceptable range, according to the BIS guideline. Water quality parameters that customers get concerned about are those whose values (derived from an average of 710 samples, split into groups of 71 samples each) fall outside of the acceptable minimum or exceed the maximum allowable value as specified. One thing to keep in mind is that different parameters have different average distances from the lower limit. If we assume that a parameter's value is better when it's closer to the acceptance threshold, then it's clear that different quality parameters have different goodness of quality. Because of this, it's reasonable to consider haracterizing water quality as a whole by establishing relationships between the goodness of separate criteria. By the way, everyone agrees that a pH of 7 is ideal for drinking water, thus pH is somewhat of an outlier in this regard.

INDEX FOR WATER QUALITY

To quantify the extent to which drinking water in the Jodhpur area is fit for human consumption, the current proposal is to create a water quality index that runs from zero to one hundred. In the BIS standard, a parameter can only have a value of 100 if its average value is equal to the specified lower limit; this is true even if the possible betterness is lower than the set lower limit. A closer look at the observed data shows that setting a lower restriction below the so-called legal limit is not really feasible.

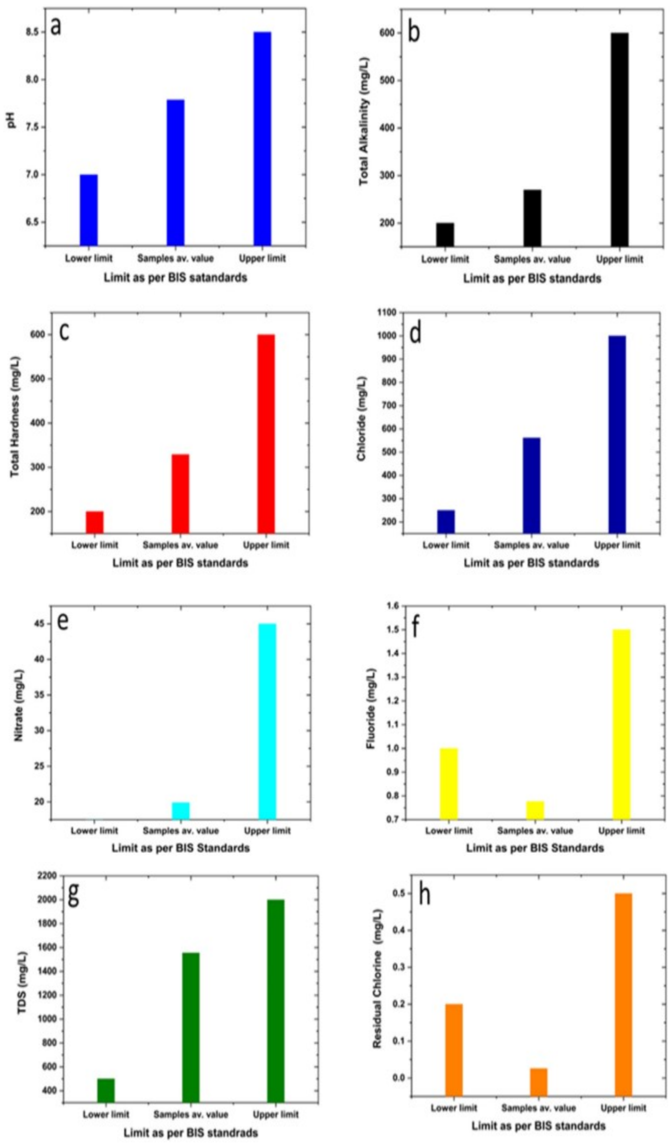


Figure 1: Average Values of Quality Parameters Over the Entire Experimental Region against its Acceptable Limits as per BIS Standards.

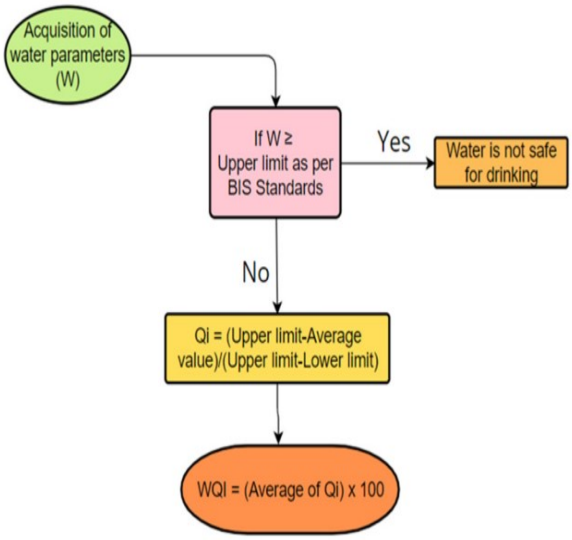


Figure 2: Flow Chart to Determine the Water Quality Index.

Research questions pertaining to the way a WQI is defined by considering the quality goodness of separate metrics are significant. It is clear that the final WQI value is determined in part by the quality index (Qi) of each parameter. Within this context, we have taken into account eight quality parameters, with 710 sampled measurement data utilised for each parameter to compute WQI in the way described above. The final water quality index is affected in diverse ways by various quality criteria. Therefore, the sum of these eight quality factors is what ultimately determines the water's quality. All eight of these quality metrics (pH, alkalinity, TDS, etc.) probably have a nonlinear connection to the so-called WQI. What this implies is that the eight quality criteria listed above form an unknown function of the WQI. To be clear, WQI might be determined using a variety of different criteria. Comparable to the BIS standard, USEPA lays forth the permissible limitations. Nevertheless, it should be mentioned that parameter specifications range among national standards; the World Health Organization (WHO) recommends standardised, rationalised approaches. Since the study is about standard-based parametric limitations, the quality index value should shift; thus, the acceptability limit should be rethought to fit the specification. Modelling the link between the eight quality variables as input and the resulting WQI as output looks intriguing, considering the volatility in values of a quality parameter with regard to time and space. There are a number of statistical methods for uncovering the causal link between an input variable and its output, but one of the most effective is the artificial neural network. Traditional perceptron training makes use of the feed-forward backpropagation method. After operating a preselected approximator (the transfer function) on weighted inputs with a bias value, the computed output is compared to the intended output.

Each iteration's observed error may be minimised by adjusting the weight value. Using this knowledge, we mapped the nonlinear relationship between the input water quality metrics and the output individual quality index using artificial neural network modelling. In addition, ANNs need the employment of an approximator, called a transfer function, in all cases involving relational pattern recognition. Any of the above transfer functions is accessible when working with the associated toolbox in MATLAB.

It is possible to employ a higher-order universal approximator as a transfer function; however, this approach necessitates the development of appropriate code for the backpropagation method as well as its integration with the associated software. Nevertheless, this will constitute independent study with little assurance of satisfactory consensus on the issue at hand. Due to the sensitive nature of the material involved, this is an exercise in informed guesswork to determine the optimal strategy. The authors disclaim any claim that their chosen approach is optimal for water quality modelling with the available data. The goal here is to guarantee an ANN's predictive ability in the simplest method possible. For the network's training, the MATLAB platform used the feed-forward backpropagation LM scheme.

The overarching data flow mechanism is identical to that of a traditional neural network. On the other hand, current ANN modelling does not make use of any prior information on the impact of the specific parameter. The goal of the supervised learning process was to uncover a hidden relational pattern among the quality indices that can be assigned to each individual parameter.

This is a first of its kind exercise because no previous example has integrated the quality indices of eight parameters using a well-known learning method. While the authors did experiment with different backpropagation algorithms in the MATLAB toolbox, the LM technique yielded somewhat better correlation. Additionally, the scaled conjugate gradient (SCG) method and gradient descent were also attempted, but failed.

The option to use solely MATLAB-compatible algorithms is at our disposal. Learning curves may get stuck in local optima for certain algorithms more often than others. It takes a variety of approaches to reach a global optimum. Using a genetic algorithm in conjunction with multivariate analysis is one such approach, as mentioned before. Alternately, one may try out unsupervised learning using a Kohonen network or neuro-fuzzy approaches, both of which contribute to effective learning. The authors have picked the easiest activity to get a feel for the interplay between the individual quality indices, but individual activity is still a big job in and of itself. Additionally, one may try using a decision support system, autoregressive moving average, or a Bayesian neural network; there are a number of additional deep learning methods that could be explored for improved prediction.

Nevertheless, the performance is deemed adequate for the ANN used in this case based on MSE, R-value, and R-square. When evaluating the performance of other processes like K-nearest neighbour, KNN, SVM, or the naïve Bayes model, additional metrics like accuracy, sensitivity, specificity, and F-score are used. In addition to MATLAB, additional useful tools for the neural network includes Tflearn, Neural designer, Keras, Neuro Solution, Torch, and Microsoft Cognitive Toolkit. To facilitate applications powered by artificial intelligence (AI), the neural designer may be used to theoretically model a comparable dataset without the need for coding. When training the network, form made use of the feed-forward backpropagation LM method.

This study follows the same basic data flow architecture as a traditional neural network, as When training reaches an acceptable low training error, the forward-flowing data is used to adjust the weights at each node using the backpropagation process, which in turn modifies the output.

It is clear that the weighted input is multiplied by a randomly selected bias value at each hidden layer node before being passed on to the transfer function, the approximator utilised in this instance. One well-known method for efficiently capturing nonlinearity is the Tanh transfer function [25,26].

On the other hand, current ANN modelling does not make use of any prior information on the impact of the specific parameter. The goal of the supervised learning process was to uncover a hidden relational pattern among the quality indices that can be assigned to each individual parameter.

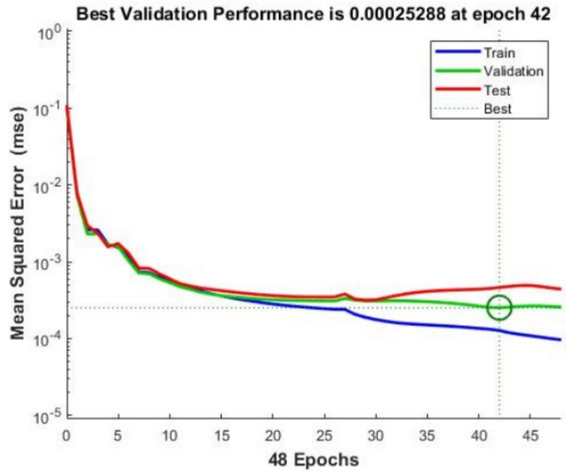
This is a first of its kind exercise because no previous example has integrated the quality indices of eight parameters using a well-known learning method. While the authors did experiment with different backpropagation algorithms in the MATLAB toolbox, the LM technique yielded somewhat better correlation. Additionally, the scaled conjugate gradient (SCG) method and gradient descent were also attempted, but failed. The option to use solely MATLAB-compatible algorithms is at our disposal. Learning curves may get stuck in local optima for certain algorithms more often than others. It takes a variety of approaches to reach a global optimum. Using a genetic algorithm in conjunction with multivariate analysis is one such approach, as mentioned before.

Alternately, one may try out unsupervised learning using a Kohonen network or neuro-fuzzy approaches, both of which contribute to effective learning. The authors have picked the easiest activity to get a feel for the interplay between the individual quality indices, but individual activity is still a big job in and of itself. Additionally, one may try using a decision support system, autoregressive moving average, or a Bayesian neural network; there are a number of additional deep learning methods that could be explored for improved prediction.

Nevertheless, the performance is deemed adequate for the ANN used in this case based on MSE, R-value, and R-square. The performance of various processes, including K-nearest neighbour, KNN, SVM, or the naïve Bayes model, is evaluated using additional metrics like accuracy, sensitivity, specificity, and Fscore. In addition to MATLAB, additional useful tools for the neural network includes Tflearn, Neural designer, Keras, Neuro Solution, Torch, and Microsoft Cognitive Toolkit. Using the neural designer, one may theoretically build a code-f dataset that is comparable to the original.

For each input variable, all 710 measurement data points were taken into account throughout the modelling process. The neural network was trained using 70% of the data, with 15% going into testing and another 15% going into model validation. Data for training, testing, and validation was chosen at random since the ANN modelling was done on the MATLAB platform.

Further randomisation of the provided data set was included in the code to support the finding from the ANN modelling in MATLAB. Multiple iterations of this random data selection process were carried out, with training, testing, and validation carried out for each of the randomly chosen data sets. Use cases for Keras include recurrent neural networks (RNNs) and convolutional neural networks (CNNs). In essence, they are deep learning programs that might be trained to solve the current challenge. In order to get a deeper understanding of the issue at hand, the writers considered using the deep learning software. This study is an exploratory analysis of the potential for mapping the relationship characteristics among the different parameters utilising learning approaches that mimic the human brain. It would not be out of place to mention the refined study by Kouadri et al. [27],



(a)

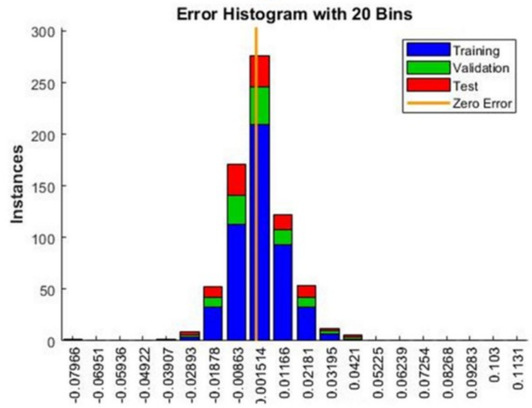


Figure 3. (a) Performance Plot of the Implemented ANN Model. (b) Error Histogram for Training, Testing, and Validation.

Table 1: 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

LIMITATIONS OF THE STUDY AND FUTURE SCOPE

The input data used to train ML models determines the models' outputs. Data used to train ML models should be big enough and varied enough to capture the intrinsic unpredictability of the underlying physical process if we want to achieve better predicting performance.

One way to handle this is by employing seasonal models or ALL-season data for model construction. This was shown in this research by identifying the seasonal effect on the RF models. As a result, you should check how much data is dependent on the ML models' prediction abilities and fix any problems you find. The ML models may also have limitations due to their site-specific nature. There was no evaluation of the created models' prediction ability in a different location as this research only looked at one site. Nevertheless, it is important to recognise that ML models built using data from several sites may not be able to generalise to other regions or that ML models built with site-specific data may lose part of their accuracy. It may be feasible to transfer the ML model from one site to another for DWQI prediction if the hydrogeological characteristics and processes in the various locations are comparable.

Such studies are suggested as a potential avenue for future study and might be carried out in various hydrogeological contexts. Water quality predictions made using hybrid or ensemble ML models have been shown to be more accurate in the literature. To improve generalisability and get beyond current model restrictions (such as the RF models' reliance on seasonal data in this research), hybrid ML models could be investigated in the future. The black-box character of ML models is one of its features; application-specific models are often chosen without providing a thorough explanation of the variables influencing the models' correctness.

For research evaluating water quality, techniques such the Shapley additive explanations (SHAP) have been used to solve this problem (Makumbura et al. 2024). This highlights the need for further study into using a similar approach to ML models created for DWQI prediction.

CONCLUSION

- The most important takeaways from the research are these: It's fair to say that both MLR and RF can model and forecast DWQI from field data; however, MLR provides better generalisation.
- Regardless of the season, the MLR models produced in this work can be used to forecast DWQI scores. This means that data from any season may be used to construct the models, and the models can be utilised for any season's predictions. On the other hand, ALL-season data can also be used, however RF models would benefit most from seasonal data.
- A correlation of 1.00 and a median error of less than 10c5 would indicate that MLR models are accurate predictors. A correlation greater than 0.99 and a median error of around 5.5 would be the outcomes of using RF models for accuracy.
- Among the ML techniques that have been shown effective in DWQI prediction, this research included MLR and RF.
- Tools for water monitoring and water quality prediction powered by AI and ML may help build a framework for managing water resources that is both sustainable and resistant to climate change.

REFERENCES

1. *Adimalla, N. & Qian, H. (2019). Spatial distribution and health risk assessment of fluoride contamination in groundwater of Telangana: a state-of-the-art, Geochemistry, 80 (4), 125548. https://doi.org/10.1016/j.chemer.2019.125548.*
2. *Akinepalli, P., Badam, A., Vithpala, P. & Bandaru, A. K. (2023) Prevalence of thyroid disorders in and around Nalgonda District, Telangana, India: a cross-sectional study, National Journal of Laboratory Medicine, 12 (1), BO20–BO24. https://doi.org/10.7860/NJLM/2023/ 59070.2694.*

3. *Alsulamy, S., Kumar, V., Kisi, O., Kedam, N. & Rathnayake, N. (2025) Enhancing water level prediction using ensemble machine learning models: a comparative analysis, Water Resources Management, 39, 3995–4014. https://doi.org/10.1007/s11269-025-04142-5.*

4. *Askari, M., Soleimani, H., Nalosi, K. B., Saeedi, R., Abolli, S., Ghani, M., Abtahi, M. & Alimohammadi, M. (2024) Bottled water safety evaluation: a comprehensive health risk assessment of oral exposure to heavy metals through deterministic and probabilistic approaches by Monte Carlo simulation, Food and Chemical Toxicology, 185, 114492. https://doi.org/10.1016/j.fct.2024.114492.*

5. *Belina, Y., Kebede, A. & Masinde, M. (2024) Comparative analysis of HEC-HMS and machine learning models for rainfall-runoff prediction in the upper Baro watershed, Ethiopia, Hydrology Research, 55 (9), 873. https://doi.org/10.2166/nh.2024.032. Bhagyalakshmi, R., Reddy, I. P., Edukondal, A., Paramesh, C., Ramu, M. & Sreenu, K. (2023) Groundwater quality and its suitability for drinking and agricultural use in part of Brahmanavellamla watershed, Nalgonda district – a case study, Journal of Applied Geochemistry, 25 (4), 227–245.*

6. *Bilali, A. E., Taleb, A. & Brouziyne, Y. (2021) Groundwater quality forecasting using machine learning algorithms for irrigation purposes, Agricultural Water Management, 245, 106625. https://doi.org/10.1016/j.agwat.2020.106625. Birkel, C., Miller, J., Watson, A., Trinh, D. A., Duran-Quesada, A. M., Sanchez-Murillo, R., Soulsby, C., Terzer-Wassmuth, S., Tetzlaff, D., Uhlenbrook, A., Vystavna, Y. & Yoshimura, K. (2025) Demystifying the art of isotope-enabled hydrological and climate modelling, Science of the Total Environment, 959, 178242. https://doi.org/10.1016/j.scitotenv.2024.178242.*

7. *Bisht, A. K., Singh, R., Bhatt, A. & Bhutiani, R. (2018) Development of an automated water quality classification model for the River Ganga. In: Bhattacharyya, P., Sastry, H., Marriboyina, V. & Sharma, R. (eds.) NGCT 2017, CCIS 827. Singapore: Springer Nature, pp. 190–198.*

8. *Breiman, L. (2001) Random forests, Machine Learning, 45, 5–32. https://doi.org/10.1023/A:1010933404324.*

9. *Brown, R. M., McClelland, N. I., Deininger, R. A. & Tozer, R. G. (1970) A water quality index – do we dare? Water and Sewage Works, (October), 339–343.*

10. *Central Ground Water Board (CGWB) (2013) Ground Water Brochure, Nalgonda District, Andhra Pradesh. Ministry of Water Resources, River Development & Ganga Rejuvenation, Government of India, Southern Region, Hyderabad, India. India: CGWB.*

11. *Central Ground Water Board (CGWB) (2016) Ground Water Brochure, Nalgonda District, Telangana State (Unpublished Report). Ministry of Water Resources, River Development & Ganga Rejuvenation, Government of India, Southern Region, Hyderabad, India. India: CGWB.*

12. *Cheng, B., Lu, K., Li, J., Chen, H., Luo, X. & Shafique, M. (2022) Comprehensive assessment of embodied environmental impacts of buildings using normalized environmental impact factors, Journal of Cleaner Production, 334, 130083. https://doi.org/10.1016/j.jclepro.2021. 130083.*

13. Dauji, S. (2023a) Climate change indications in short-period rainfall records from coastal India using trend analysis: a proposed framework, Arabian Journal of Geosciences, 16, 189. https://doi.org/10.1007/s12517-023-11269-5.

14. Dauji, S. (2023b) Comparison of NDT data fusion for concrete strength using decision tree and artificial neural network, Journal of Scientific & Industrial Research, 82, 831–840. https://doi.org/10.56042/jsir.v82i08.3048.

15. Dauji, S. (2024a) Probabilistic effect of non-potable mixing water on characteristic strength of concrete, Magazine of Concrete Research, 76 (13), 707–722. https://doi.org/10.1680/jmacr.23.00133.

16. Dauji, S. (2024b) Probabilistic acceptance criteria for strength of concrete produced using non-potable water for sustainable infrastructure, Construction and Building Materials, 447, 137706. https://doi.org/10.1016/j.conbuildmat.2024.137706.

17. Dauji, S., Pant, D. & Tirumalesh, K. (2023) Probabilistic assessment of groundwater quality and fluoride exposure risk for North-Eastern Rajasthan, India, Journal of Hazardous, Toxic, and Radioactive Waste, 27 (4), 04023030. https://doi.org/10.1061/JHTRBP.HZENG-1254.

18. Dharmarathne, G., Abekoon, A. M. S. R., Bogahawaththa, M., Alawatugoda, J. & Meddage, D. P. P. (2025) A review of machine learning and internet-of-things on the water quality assessment: methods, applications and future trends, Results in Engineering, 26, 105182. https:// doi.org/10.1016/j.rineng.2025.105182. Editorial (2021) Concrete needs to lose its colossal carbon footprint, Nature, 597, 593–594.

19. Essamlali, I., Nhalia, H. & Khaili, M. E. (2024) Advances in machine learning and IoT for water quality monitoring: a comprehensive review, Heliyon, 10, e27920. https://doi.org/10.1016/j.heliyon.2024.e27920. Fawcett, T. (2006) An introduction to ROC analysis, Pattern Recognition Letters, 27, 861–874. https://doi.org/10.1016/j.patrec.2005.10.010.

20. Freedman, J. & Tayalia, Y. (2020) Addressing India's Water Crisis – Policy Options to Promote Water Recycling and Reuse. Paris: SUEZ Water Technologies and Solutions. Available at: https://www.suez.com/-/media/suez-global/files/india/publication-docs/addressingindia-water-crisis-through-reuse-2020.pdf (Accessed: 19 January 2023).

21. Ganaboina, G. M. & Kota, N. (2021) A case study from Ramannapeta Mandal, Nalgonda, Telangana, India: fluoride contamination of ground water, Preprint: Research Square. https://doi.org/10.21203/rs.3.rs-681568/v1.

22. Gichamo, T., Nourani, V., Gokcekus, H. & Gelete, G. (2024) Ensemble of artificial intelligence and physically based models for rainfall– runoff modeling in the upper Blue Nile Basin, Hydrology Research, 55 (10), 976. https://doi.org/10.2166/nh.2024.189.

23. Goodarzi, M. R., Bafrouei, H. B. & Vazirian, M. (2025) Insight into groundwater level prediction with feature effectiveness: comparison of machine learning and numerical methods, Hydrology Research, 56 (1), 74. https://doi.org/10.2166/nh.2024.135.

24. Ground Water Department (GWD) (2021) Seasonal Groundwater Quality Report during Pre and Post-Monsoon Season. Unpublished Data. File No. 6479/QW Level II þ Lab/2021, GWD, Government of Telangana, Telangana, India. India: GWB.

25. *Guguloth, S., Pandey, M. & Pal, M. (2024) Application of hybrid AI models for accurate prediction of scour depths under submerged circular vertical jet, Journal of Hydrologic Engineering, 29 (3), 04024010. https://doi.org/10.1061/JHYEFF.HEENG-6149.*
26. *Guhathakurta, P., Sreejith, O. P. & Menon, P. A. (2011) Impact of climate change on extreme rainfall events and flood risk in India, Journal of Earth System Science, 120, 359–373. https://doi.org/10.1007/s12040-011-0082-5.*
27. *Hassan, M. M., Hassan, M. M., Akter, L., Rahman, M. M., Zaman, S., Hasib, K. M., Jahan, N., Smrity, R. N., Farhana, J., Raihan, M. & Mollick, S. (2021) Efficient prediction of water quality index (WQI) using machine learning algorithms, Human-Centric Intelligent Systems, 1 (3–4), 86–97. https://doi.org/10.2991/hcis.k.211203.001.*