SPC Journal of Environmental Sciences, 8 (1) (2026) 1-10

SPC Journal of Environmental Sciences

Website: www.sciencepubco.com/index.php/JES
https://doi.org/10.14419/hs5ywd96
Research paper

A Comparative Analysis of Groundwater Quality and Its Prediction Using Machine Learning Techniques

B. Vamsi [1], B. Chandra Sekhar [2] *, B. Praveen Kumar [3], G. Rajling [4], K. Satya Gayathri [5], N. Mahendra [5], O. Sudha [5], K. Bobby [6], P. V. S. S. Manikanta Suhas [5], K. Surya Prakash [7], Ch. Srinivas [5], M. Gopi Kiran [1]

[1] Department of Civil Engineering, Lingayas Institute of Management and Technology, Madalavarigudem, Vijayawada-521212, India
[2] Department of Chemical Engineering, RGUKT RK Valley, Iddupulapaya, Vempalli, YSR Kadapa-516330, India
[3] Research Scholar, Department of Geology, Andhra University-530003, Andhra Pradesh, India
[4] Arc Infra, Civil Engineering Consultant, Vijayawada, Andhra Pradesh 520008, India
[5] Department of Computer Science and Engineering, Lingayas Institute of Management and Technology, Madalavarigudem, Vijayawada-521212, India
[6] Department of Artificial Intelligence and Machine Learning, Lingayas Institute of Management and Technology, Madalavarigudem, Vijayawada-521212, India
[7] Department of Artificial Intelligence and Data Science, Lingayas Institute of Management and Technology, Madalavarigudem, Vijayawada-521212, India
**Corresponding author E-mail: drbandi85@rguktrkv.ac.in*

Received: November 13, 2025, Accepted: January 29, 2026, Published: February 10, 2026

Abstract

Groundwater is a key water source; even small variations in its quality may result in detrimental impacts, necessitating its regular monitoring for effective water management. This study emphasized monitoring and assessing groundwater from Mudirajupalem, Madalvarigudem, and the Lingayas Institute of Management and Technology, Vijayawada campus, Andhra Pradesh, India. The groundwater was analyzed for quantifying acidity, alkalinity, pH, total dissolved solids (TDS), and total hardness according to standard methods. All the groundwater sources showed high alkalinity values, whereas Mudirajupalem groundwater exhibited high TDS (962±20 mg/L) and pH (9.2±0.3) values due to proximity to farming fields. The obtained water quality index (WQI) values (358.34±15.2, 59.29±4.8, and 20.32±3.1 for the three sites, respectively) indicate that Mudirajupalem groundwater is unfit for public intake. This study demonstrated promising results in forecasting water quality parameters using a hierarchical reconciliation algorithm and predicted the WQI using gradient boosted tree (GBT), random forest (RF), and decision tree (DT) techniques. The predicted WQI values closely matched the experimental results, confirming that Mudirajupalem groundwater is not fit for drinking. The GBT demonstrated superior performance (R^2=0.95-0.98) compared with RF (R^2=0.89-0.93) and DT (R^2=0.84-0.87) for the selected study area , and this study demonstrates that the application of advanced machine learning enables a proactive approach for better water quality management to address the future water needs.

***Keywords**: Use about five keywords or phrases in alphabetical order, separated by a semicolon.*

1. Introduction

Water is very much essential for life on Earth. Groundwater serves as a primary water source in arid and semi-arid regions, gaining prominence due to increasing population growth and several other demands. [1]. The central groundwater board categorizes groundwater as semi-confined or unconfined, existing in sandy clays, gravels, and sands. [2]. Water availability directly influences a nation's economic and social development, making groundwater management critical for sustainable development. [3]. Population growth has dramatically increased demand for freshwater resources, wherein a primary solution is groundwater apart from rainwater and other sources. However, the global water supply is majorly influenced by ever-growing concerns of climate change, owing to human-induced developmental activities. [4]. Nevertheless, the natural existence of salinity in the groundwater limits its applicability for various needs. Therefore, it is indisputably essential to sustain the potability of water for human existence.

In most of the places around the globe, one of the water sources free from any sort of pollution and clean is groundwater, matching surface water. In a study, Gleeson et al. [5] reported that approximately two billion people depend on groundwater as a potable water source, whereas almost 40% of irrigation needs are met by groundwater globally. Several developing countries, such as Iran, India, etc., are facing comparatively limited groundwater resources, owing to the issues associated with human existence as a result of depleting groundwater resources and declining water quality owing to numerous causes [6-8]. The recent investigations on groundwater quality for irrigation reported that there is a rise in groundwater depletion owing to escalating population, developmental activities, changes in the agricultural and waste disposal practices [9]. Sustainable Development Goal (SDG) 6 is marked as a key asset to make water management sustainable and sanitation available by 2030 as per the United Nations SDGs [10]. The current level of groundwater diminution is one of the critical issues, making it a challenge in accomplishing the SDGs [11].

Numerous tools and techniques, such as the water quality index (WQI), analytic hierarchy processes, fuzzy mathematics, neural networks, etc., have been employed to assess the quality of groundwater. [12-16]. Hence, regular monitoring of groundwater resources ensures water safety for human use. Machine learning (ML) technologies have attained wide acclaim for prediction and their applicability in various fields. Inevitably, ML techniques can also be employed for the prediction of groundwater availability for effective water management. The ML techniques enable the collection of datasets, storing them in databases, and processing to develop worthwhile insights, which can be used for various applications. [17]. In recent years, numerous investigations have been performed to forecast groundwater quality using ML techniques, such as random forest (RF) [18], multiple linear regression (MLR), decision tree regression (DTR), RF regression (RFR), and artificial neural network (ANN) [19], etc., which proved to be potential and reliable in successfully predicting the groundwater quality. Therefore, it is vital to monitor the groundwater resources for their quality and to manage the water resources systematically, and make use of modern methods for groundwater quality forecasting. Therefore, the objective of the present study mainly focused on monitoring and assessing the deviation in the groundwater quality from three different sources, as there is a remarkable gap in the groundwater quality data in the selected region. The sources include Mudirajupalem and Madalvarigudem, Krishna district, Andhra Pradesh, India, and Lingayas Institute of Management and Technology (LIMAT), Vijayawada campus, Andhra Pradesh, India. The groundwater analysis was carried out for the parameters of acidity, total dissolved solids (TDS), total hardness (TH), alkalinity, and pH as described in standard methods, as these are a few commonly analyzed parameters for water analysis. Also, as an initiation to the implementation of modern ML techniques, this study employed a few ML techniques, such as decision tree (DT), RF, and gradient boosted trees (GBT) for predicting the groundwater quality, as these are popular and well-employed methods in the literature for various applications.

This study integrates conventional methods of groundwater quality analysis with ML techniques, such as DT, RF, and GBT, to predict groundwater quality in the selected study area. Additionally, this research fills a significant data gap in groundwater quality assessment for Mudirajupalem, Madalvarigudem, and LIMAT, Andhra Pradesh, India, where studies utilizing intelligent technology for groundwater monitoring remain limited. By leveraging ML techniques, this study not only provides the evaluation of groundwater quality but also introduces predictive models that can serve as decision support tools for effective water resource management. This approach is expected to be a stepping-stone for broader applications of ML in hydrology research in the selected study area and contribute to achieving SDG 6. This study presents findings from an academic institution within the study area, chiefly underlining the magnitude of the quality of groundwater in terms of WQI and predicting its accuracy. This could eventually be beneficial to the public, as water is a valuable resource of nature, which would further delve into interpreting the quality of groundwater of the study area for various purposes in the future. Moreover, the present study lays down a durable platform to perform productive research on groundwater quality prediction and proposes relevant locale-specific techniques to conserve water resources that cater to the future water needs and enhance the water quality for public use.

2. Materials and Methodology

To plan and execute the monitoring of groundwater quality, samples were collected from three sources, such as Mudirajupalem and Madalvarigudem, Krishna district, Andhra Pradesh, India, and LIMAT, Vijayawada campus, Andhra Pradesh, India. Groundwater quality monitoring was performed by collecting groundwater samples for a period of more than one month, i.e., from March to April 2024, keeping the time frame and limited resources available in view. The groundwater analysis was carried out to quantify the acidity, TH, pH, TDS, and alkalinity by making use of the standard methods described in the American Public Health Association (APHA) [20].

The LIMAT campus is situated in Madalavarigudem village, pertaining to the Gannavaram Mandal of Krishna District, Andhra Pradesh, India. Agiripalli Mandal borders Madalavarigudem on the way to the north, Unguturu Mandal in the direction of East, Kankipadu Mandal, and Penamaluru Mandal to the South. The satellite images of the groundwater sources in the selected study area are shown in Figure 1. The TH, acidity, and alkalinity in all the samples were quantified by making use of titration methods as per the APHA standard methods. Scientific pH papers (S. D. Fine-chem Ltd, Mumbai, India) were used to measure the samples' pH collected from three sources. The TDS in all the samples was measured using a Generic Digital liquid crystal display TDS Meter (Generic, TDS-3, India). The World Health Organization (WHO) and the Bureau of Indian Standards (BIS) [21] were taken as a baseline for assessing the water quality. The method described by Brown et al. [22] was used to calculate the WQI of the samples to examine the feasibility of the groundwater quality for numerous applications.

(A)

(B)

(C)

Fig. 1: Satellite Images of the Selected Study Area A.) the LIMAT, Vijayawada Campus, B.) Mudirajupalem, and C.) Madalvarigudem, Krishna District.

This study made use of ML techniques, such as DT, RF, and GBT, for forecasting the groundwater quality in the selected study area, as these are widely employed methods in the literature for various applications. The RF is a well-employed ML technique, which conglomerates the forecasts of multiple decision trees for performance enhancement. Reduced variance, high accuracy in predictions, and less susceptibility to overfitting are some of the pros of the RF technique, apart from making it suitable for larger datasets with various features. [18]. In a DT technique, any trail opening from the root is designated as a data splitting system until a Boolean outcome is achieved at the leaf node. [23-25].

The GBT uses decision trees for predictions, in which the weaker estimation method is commonly used. It is very efficient at forecasting as it combines several smaller, inefficient models into one robust model, which works effectively. This method is very popular as it is efficient at rapidly classifying datasets. [26]. All models were implemented using TensorFlow Decision Forests (Google Zurich), an open-source framework facilitating consistent preprocessing and validation across algorithms. Following established ML protocols [27, 28], the dataset was partitioned using stratified random sampling: 70% for training (n=74 samples) and 30% for testing (n=31 samples), ensuring representative class distribution in both subsets. This study implemented a multi-level validation framework combining three complementary approaches to ensure robust performance estimates [27, 28] and time series-specific cross-validation [29]. Also, a hierarchical reconciliation algorithm was used to forecast the future trend of the parameters tested for quantifying the WQI values [30].

3. Results and Discussion

Groundwater accounts for half of Indian urban water usage. However, proper light is not shed on its sustainable usage, as checking and assessment of water quality is limited on a large scale. To develop a solid solution for water-based environmental setbacks, it is essential to consistently monitor all water resources to collect accurate information, identify issues, and understand their status and root causes. This would further aid in devising a reasonable and robust way out to water related problems. One of the limitations of the widely used parameters to figure out the intensity of water quality is TDS, which is characterized by the dissolved materials existing in the water. Figure 2 depicts the difference in the TDS value of the groundwater samples. Mudirajupalem exhibited the highest TDS values among the tested sites, which ranged between 931-994 mg/L, followed by LIMAT campus (199-273 mg/L) and Madalavarigudem (117-280 mg/L). Normally, oozing of salts from soil and percolation of sewage from domestic sources may result in increasing the TDS values. [31].

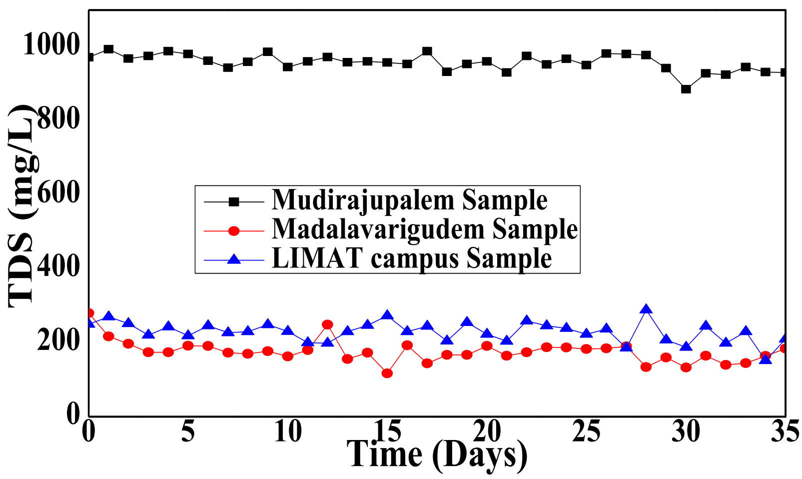

Fig. 2: The Difference in the TDS Values in the Samples of Groundwater of the Study Area.

High TDS values in the Mudirajupalem samples are likely attributed to the existence of higher concentrations of both dissolved and suspended substances added by human-induced activities. This is also true in line with the results of interviews and surveys conducted in the present study with the local public in the selected location, also in agreement with the results reported in the literature. [32]. The values of TDS in Madalavarigudem and LIMAT campus groundwater were below the allowable TDS limit, which is described in the BIS standards; however, it is indispensable to treat the groundwater to make it fit for human consumption. Treatment technologies, including reverse osmosis (RO), deionization (DI), and distillation, effectively reduce TDS to acceptable levels.

Figure 3 illustrates the difference in the pH values in the groundwater samples of the study area. The strength of hydrogen ions present in water is reflected by the pH value, which stimulates reactions in any aquatic body; thereby, it is regarded as a chief feature in aquatic systems. [33-34]. Among samples tested during the monitoring period, Mudirajupalem groundwater showed maximum pH values that ranged from 8.5-10, signifying that the pH values are slightly more than the allowable limit of the BIS standards. Whereas, the pH values ranged between 6.5-9.0 and 6.5-8.0 for the samples of Madalavarigudem and the samples of LIMAT campus, respectively, indicating that the pH values in the latter samples are within the allowable limits in line with the BIS standards. High pH values may lead to corrosion of pipes, unpleasant taste, etc.

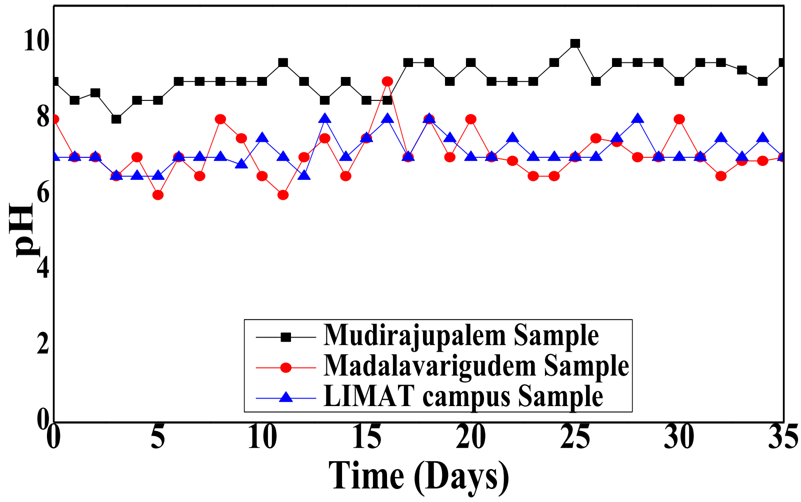

Fig. 3: Illustration of the pH Values About Three Sources of the Study Area.

Mudirajupalem groundwater samples showed higher pH values, which can be attributed to the presence of agricultural fields within the vicinity of the study area, as is also true in the context of the results reported by Khatri et al. [35]. It can be inferred from the results that the range of pH values in the three groundwater sources indicates the alkaline nature, likely due to the existence of diverse buffers in the groundwater. [36]. Whether it is chemical or physical factors, the quality of groundwater is bound to vary depending on various factors, such as geological factors, human-made actions, and climate change. [37, 38]. Besides, the chemistry of groundwater with respect to a certain rock fluctuates depending on the contact time; the longer the contact time, the more the chances of the effect of rock chemistry on pH and the quality of groundwater. Moreover, pH fluctuates with the rocks' configuration and deposits edging the path through which recharge water penetrates the groundwater. [39].

Basically, the alkalinity in water is characterized by the acid-neutralizing capacity. [40]. The difference in the alkalinity values in the samples of groundwater in the study area is presented in Figure 4. Among the samples tested, Madalavarigudem groundwater showed maximum alkalinity values, which ranged between 525-910 mg/L, whereas in the context of Mudirajupalem and the LIMAT campus groundwater, alkalinity values ranged between 460-850 mg/L and 550-750 mg/L, respectively. In general, the alkalinity of water samples is bound to vary with the amount of cations, anions, alkaline metals, etc., that exist in the natural water. [31]. The alkalinity values obtained in this study imply that the values are far more than the permissible limits described in the BIS standards. Elevated alkalinity concentrations can be reduced by employing water-softening processes.

Figure 5 highlights the difference in the acidity values in the samples of groundwater in the study area. Among the samples tested, the LIMAT campus groundwater showed maximum acidity values, which ranged between 45-525 mg/L, whereas in the context of Mudirajupalem and the Madalavarigudem groundwater, acidity values ranged between 45-480 mg/L and 105-460 mg/L, respectively. Groundwater is bound to remain acidic in nature, depending on the aquifer geology containing carbonate rocks. As reported in the literature, acidic quality in water in itself is not so harmful; perhaps, the conduits or pipelines that convey the drinking water may get affected even if the water is in mild acidic nature. Therefore, several standards, which have been framed to tune the quality of water, have determined that a pH value ranging between 6.5 and 8.5 would be a barrier to minimize the scale deposits and dissolved substances because of acidic water. [39].

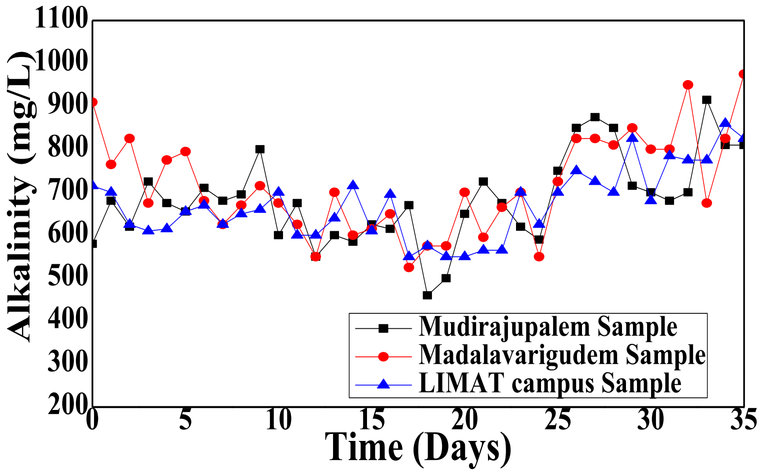

Fig. 4: Variation of the Alkalinity Values in Three Sources of the Study Area.

Hardness in water comes into the picture because of polyvalent metallic ions dissolved in water, more predominantly due to magnesium and calcium. [41]. The difference in the TH values in the samples of groundwater in the study area is presented in Figure 6. Among the samples tested, the Mudirajupalem groundwater showed maximum TH values, which ranged between 20 and 246 mg/L, whereas in the context of the LIMAT campus and the Madalavarigudem groundwater, TH values ranged between 20-230 mg/L and 10-210 mg/L, respectively. According to Sawyer [41]The relationship between TH and the nature of the water varies, as soft with TH less than 75 mg/L, moderately hard with TH ranging between 75-150 mg/L, hard with the TH ranging between 150-300 mg/L, and very hard if the TH is more than 300 mg/L. The TH values in groundwater from all three sources in this study indicate that the groundwater falls under a soft to hard band.

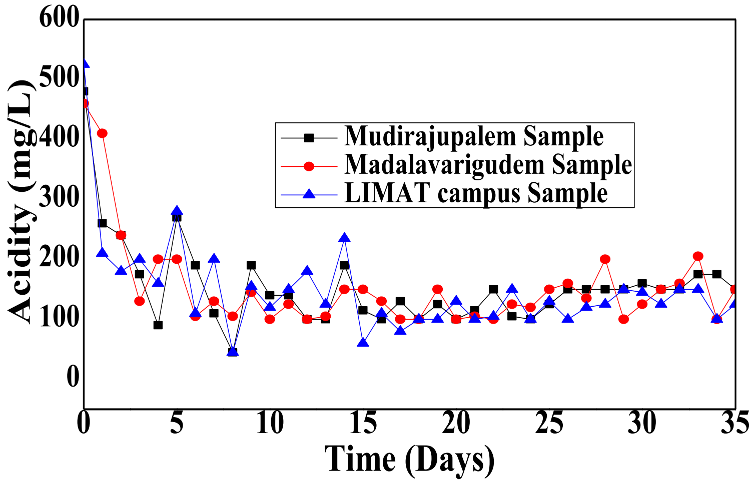

Fig. 5: Acidity Value Variation in the Samples Collected from Different Sources.

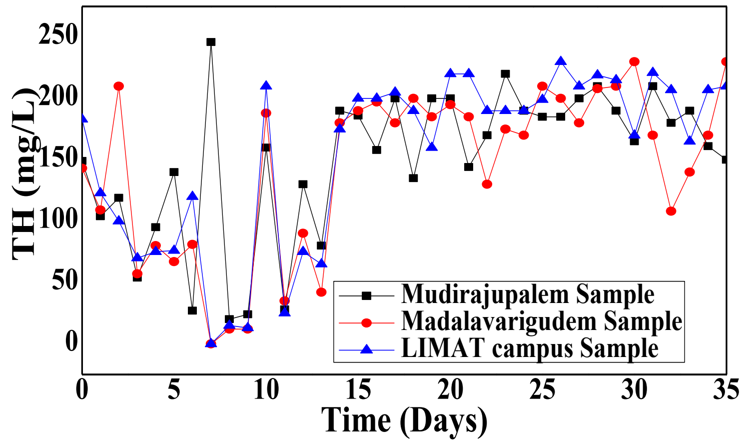

Fig. 6: Source-wise Variation in TH Values Across the Samples from Different Sources.

Alkaline earth materials, such as magnesium and calcium, impart hardness to the water. The existence of non-carbonate and carbonate compounds in water will result in raising the total hardness level. [42]. The TH values in all three sources are slightly higher than the allowable limits approved by the standards of BIS, inferring the hard water nature of the groundwater. Water softening processes are widely popular for removing excess TH concentration from water. Among several contemporary tools available for gauging the quality of water, the WQI is deliberated as a key resourceful technique used for computing the quality of water. Brown et al. [22] reported that the WQI could be used to examine fluctuations in the quality of water and can be compared with any other water supply quality. The WQI takes into account several parameters, which convert an ample amount of data to a single number that enunciates the overall water quality.

The WHO described the variation of the quality of water with WQI values, which state that water is unfit for drinking if the WQI is more than 100, very poor water quality if the WQI ranges between 76-100, poor water quality if the WQI ranges between 51-75, good water quality if the WQI ranges between 26-50, and excellent water quality if the WQI is less than 25 [43]. Amongst the samples of groundwater tested in the study area, the WQI values followed the order of 358.34±15.2, 59.29±4.8, and 20.32±3.1 for groundwater from Mudirajupalem, Madalavarigudem, and LIMAT campus, respectively, demonstrating that Mudirajupalem groundwater doesn't entail fit for public water supply. [43]. In line with the relationship between WQI and water quality, Madalavarigudem groundwater is considered to have poor water quality, whereas the LIMAT campus groundwater is of good water quality, which could be utilized for public water supply, provided after appropriate treatment for safety purposes. In a study conducted to investigate the groundwater quality and potential analysis using geospatial techniques, it was reported that the WQI values were in the range of 5.208-134.232 [44]. Similarly, a study was conducted to investigate the groundwater quality evaluation using WQI and geospatial techniques, wherein the WQI values were in the range of 77.03-115.14 [45], Both the study results slightly matching with the current study results.

The results from the groundwater analysis showed high values of alkalinity in all three sources, whereas among the three sources, Mudirajupalem groundwater exhibited high pH and TDS values, which could be due to the presence of agricultural fields within the study area. Khatri et al. [35] demonstrated comparable results in a study conducted, wherein hardness, alkalinity, and TDS concentration in the groundwater were found at elevated levels, which is owed to the geological strata. Also, there is a tendency for the increased levels of alkalinity and hardness in the groundwater because of the oozing of magnesium and calcium into the groundwater. This is true in the context of the

results obtained in the current study. Apart from this, it was very well reported in the literature that the geology of the bedrock and weather conditions play a chief role in the groundwater quality in southern India, which is influenced by human-induced activities [46].
The authors of this study conducted ground-level awareness campaigns in and around Mudirajupalem to inform the public on restricting the usage of groundwater in Mudirajupalem for drinking purposes and how it is going to affect the health, making use of information and communication technology (ICT) tools. The authors have already submitted a separate consultancy report on the findings of this study to the local Government along with simple and cost-effective designs for the implementation of rainwater harvesting techniques as well. The authors have implemented two simple and cost-effective designs of rainwater techniques in the LIMAT campus to recharge the groundwater with rainwater.
The future-forecasting trend of the selected water quality parameters was performed using a hierarchical reconciliation algorithm for nearly two months, where the forecasted TDS, pH, TH, and alkalinity values of the samples collected from three sources, along with the experimental results, are depicted in Figures 7 to 10. The reconciliation algorithm used is a top-down approach using forecast proportions. The most common top-down approaches specify disaggregation proportions based on the historical proportions of the data. As per the literature, these approaches are simple to use and produce quite reliable forecasts for the aggregate levels, and they are useful with low-count data. [30]. After incorporating the forecasted values, the WQI values of the groundwater from LIMAT campus, Madalavarigudem, and Mudirajupalem followed the order of 26.22, 63.81, and 362.96, respectively, matching the experimental results obtained, further validating that Mudirajupalem groundwater is not fit for public consumption [47]. Sensitivity analysis revealed that TDS exerts a strong influence on WQI, followed by alkalinity.

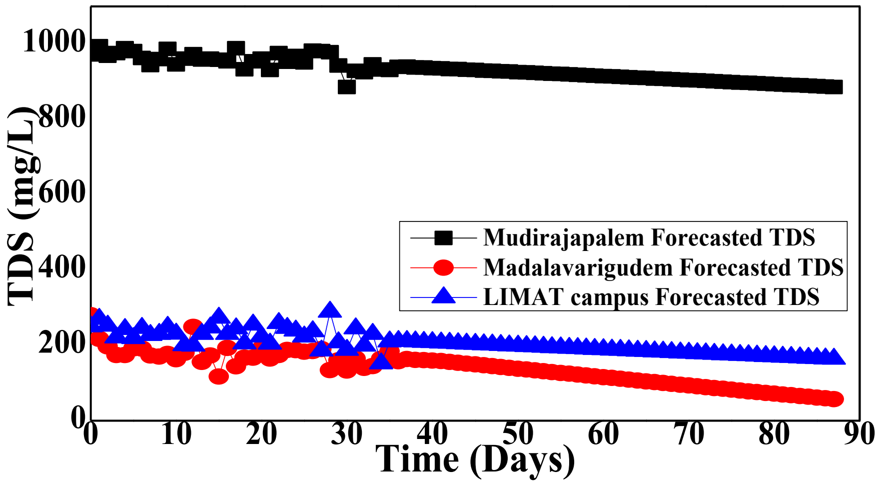

Fig. 7: Experimental Results and Forecasted Total Dissolved Solids (TDS) Values From Three Sources.

As the WQI is a numeric value, ML techniques, such as GBT, RF, and DT, were used for the groundwater quality prediction. The GBT resulted in R^2 values of 0.95, 0.98, and 0.88 for Mudirajupalem, Madalavarigudem, and LIMAT campus samples, respectively. The RF resulted in R^2 values of 0.89, 0.93, and 0.90 for Mudirajupalem, Madalavarigudem, and LIMAT campus samples, respectively, whereas DT resulted in R^2 values of 0.87, 0.84, and 0.845 for Mudirajupalem, Madalavarigudem, and LIMAT campus samples, respectively. The GBT resulted in root mean squared error (RMSE) values of 17.55, 20.66, and 66.85 for Mudirajupalem, Madalavarigudem, and LIMAT campus samples, respectively. The RF resulted in RMSE values of 17.55, 48.86, and 37.7 for Mudirajupalem, Madalavarigudem, and LIMAT campus samples, respectively, whereas DT resulted in RMSE values of 23.69, 39.32, and 25.73 for Mudirajupalem, Madalavarigudem, and LIMAT campus samples, respectively.

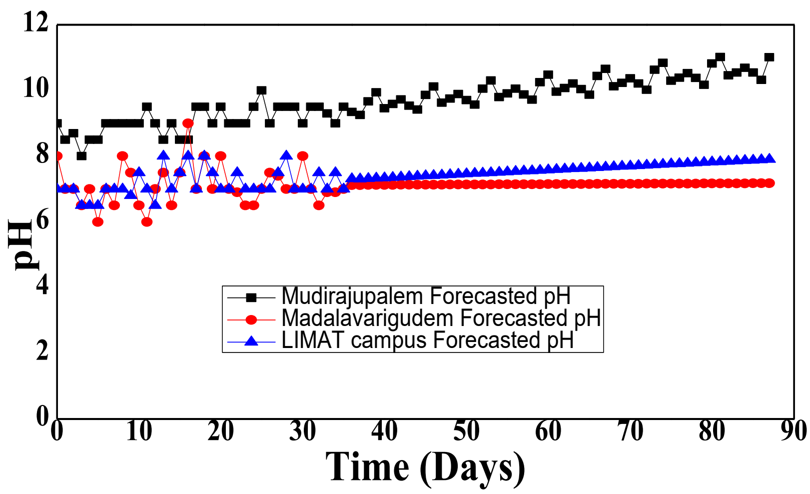

Fig. 8: The Experimental Results and Corresponding Forecasted pH Values for Selected Three Sources.

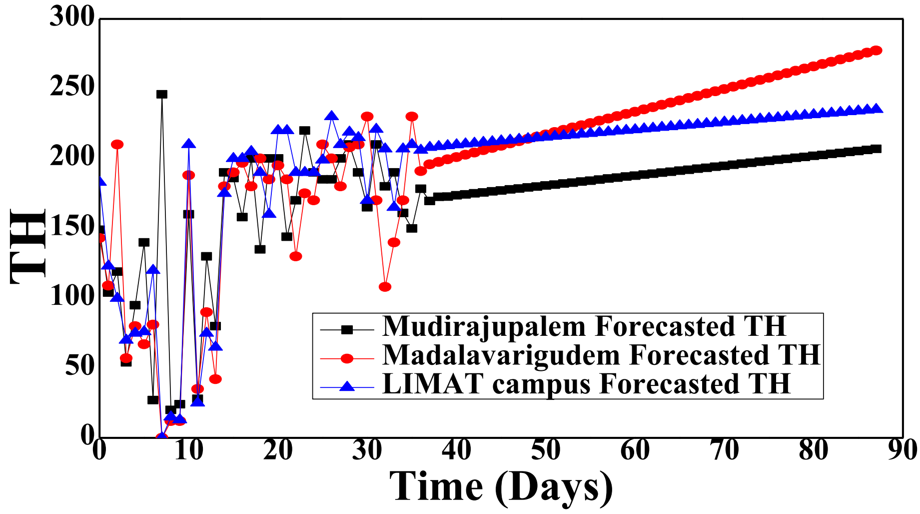

Fig. 9: Illustration of the Experimental Results and Corresponding Forecasted TH Values in Three Sources.

Apogba et al. [18] Their study yielded an RMSE of 23.03 and an R^2 value of 0.82 using RF, which is in agreement with the current study results, indicating that GBT and RF can foretell and produce favorable results. In a study planned for groundwater characterization and quality forecasting using a combination of multivariate statistics and ML techniques, DTR and RFR yielded RMSE values of 6.8 and 6.0 with R^2 values of 0.89 and 0.91, respectively [19]. Table 1 presents a comparison of groundwater quality prediction studies using various ML techniques.

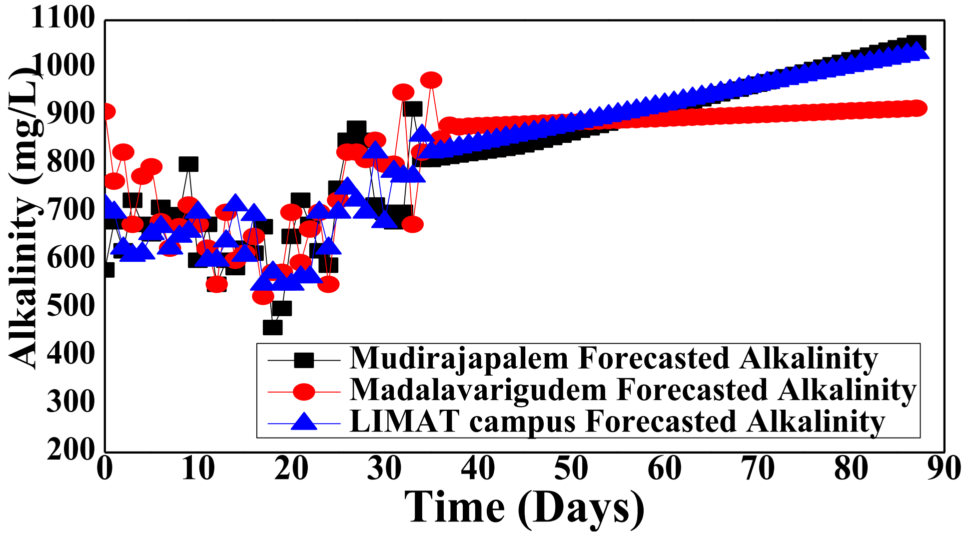

Fig. 10: Experimental and Forecasted Alkalinity Values for Groundwater Samples from Three Sources.

Table 1: Comparison of Groundwater Quality Prediction Studies Using Various ML Techniques

Study	Location	ML Techniques	Sample Size	Parameters	Performance (R^2/RMSE)	Strengths	Limitations	Compared with the present study
Apogba et al. [18]	Nabogo Basin, Ghana	RF	150 samples	12 parameters	R^2=0.82, RMSE=23.03	First ML study for the selected region. Comprehensive parameter set	No algorithm comparison. Low sample to parameter ratio (12.5:1). No temporal analysis. Missing uncertainty quantification	Higher sample density (21:1 ratio). Three-algorithm comparison. Temporal forecasting added
Abu et al. [19]	Northern Ghana	DTR, RFR, MLR, ANN	500 samples	5 parameters	RFR: R^2=0.91, RMSE=6.0; DTR: R^2=0.89, RMSE=6.8	Large dataset Multiple algorithm comparison Strong validation	Single time point No GBT evaluation Limited to retrospective analysis No prediction intervals	Adds GBT to the comparison. 35-day continuous monitoring. 60-day capability.
Ubah et al. [50]	Nigeria	ANN	120 samples	4 parameters	R^2=0.93	ANN architecture optimization Irrigation-specific	Single algorithm. No ensemble methods. Limited parameters.	Ensemble methods comparison. Five parameters.

Khatri et al. [35]	Gujarat, India	Traditional WQI only	45 samples	11 parameters	WQI calculated; ML was not applied	Regional relevance (India) Agricultural context	No validation strategy reported No predictive capability. Static assessment. ML was not implemented.	k-fold validation. Drinking water focus. ML application for a similar Indian context. Predictive modeling added. Forecasting.
Kumar et al. [45]	Andhra Pradesh, India	Traditional WQI and GIS	30 samples	8 parameters	WQI range: 77-115	Andhra Pradesh GIS integration	ML techniques were not employed. Small sample size. Single time point. No prediction	Same geographic region. ML implementation. Continuous monitoring. 60-day forecasting.
Current Study	Krishna District, Andhra Pradesh, India	GBT, RF, DT	105 samples (35 days × 3 sites)	5 parameters	GBT: R^2=0.95-0.98; RMSE=17.55-66.85	Comparison of GBT/RF/DT for groundwater WQI. Continuous 35-day monitoring. 60-day hierarchical forecasting. Agricultural contamination context	Moderate sample size. Single region. Limited to 5 parameters.	Three-algorithm comparison. Forecasting integration. ML & WQI integration

The current study results are in agreement with the literature in terms of R^2, even though the RMSE values differ. Similar results (R^2 of 0.92, 0.98, and 0.96) were reported by Wang et al. [48], Sami et al. [49], and Ubah et al. [50] in respective research works carrying out the WQI analysis. The variation in the RMSE and R^2 values could be attributed to different operating conditions, features, and the range of values in the dataset. Among the GBT, RF, and DT techniques employed, GBT and RF generated encouraging results that are identical to the experimental results [47]. Apogba et al. [18] In their study used twelve parameters with 150 samples, while Abu et al. [19] used five parameters with 500 samples (100 samples per parameter) in their investigation. The difference in sample to parameter ratio likely explains performance variation, suggesting RF requires substantial data density for optimal performance. This study addresses this by maintaining a favorable ratio (35 samples × 3 locations = 105 samples for five parameters, 21:1 ratio). Abu et al. [19] reported an R^2 value of 0.89 with an RMSE value of 6.8 for DT regression, outperforming several studies using algorithms that are more complex. However, Eesa et al. [23] and Yang [24] reported the susceptibility of DT to overfitting with noisy environmental data. Studies by Sami et al. [49] (n=1,095) and Ibrahim et al. [9] (15 parameters with GIS) establish the potential for broader approaches. This study systematically compared DT, RF, and GBT under identical conditions to figure out which model performs better performance.

Moreover, the GBT resulted in WQI values of 396.74, 33.62, and 40.05 for Mudirajupalem, Madalavarigudem, and LIMAT campus samples, respectively. The RF resulted in WQI values of 397.09, 33.66, and 39.59 for Mudirajupalem, Madalavarigudem, and LIMAT campus samples, respectively, whereas DT resulted in WQI values of 396.73, 33.59, and 41.01 for Mudirajupalem, Madalavarigudem, and LIMAT campus samples, respectively. Even though there was a variation in the WQI values of samples of Madalavarigudem and LIMAT campus, WQI values of Mudirajupalem samples are high, which are once again closely matching with the obtained experimental results, further validating that Mudirajupalem groundwater is not fit for public consumption. In a study carried out to characterize the groundwater quality and forecast, Abu et al. [19] reported WQI values ranging from 15 to 374. Similarly, in a study planned for the characterization of the groundwater quality and forecasting, Apogba et al. [18] reported WQI values ranging from 9.51 to 69.99 with different features and conditions. The variation in WQI values may be due to different hydrological conditions and features applied in different investigations. The GBT exhibited maximum R^2 values consistent with the findings of Wang et al. [48] and Sami et al. [49]. The RF exhibited moderate R^2 values consistent with the findings of Abu et al. [19], followed by DT, which exhibited slightly lower R^2 values. The findings of this study highlight the significance of context-specific model development rather than algorithm generalization across different hydrogeological conditions.

This study substantiates the necessity to perform productive research on groundwater quality analysis with additional water quality parameters for a long period covering seasonal variations to generate real time database in the selected study area, eventually benefiting society. Also, extend this study to employ advanced ML techniques for predicting more accurate groundwater quality assessment. This study's results also envisage that it is essential to embrace location-specific treatment techniques to make the groundwater safe for human sustenance. Also, this study foresees the necessity of rainwater harvesting structures implementation to increase the groundwater recharge so as to meet the future water demand, as groundwater is undeniably one of the key sources of water supply globally.

4. Conclusion

Overall, this study addressed the monitoring and evaluation of the quality of groundwater from three different sources for a period of more than one month, analyzing TH, TDS, pH, acidity, and alkalinity. Groundwater analysis revealed high values of alkalinity in all three sources; however, Mudirajupalem exhibited elevated pH (9.2±0.3) and TDS (962±20 mg/L), which could be due to the presence of nearby agricultural fields. The WQI values of the samples tested in the study area demonstrated that Mudirajupalem groundwater is not appropriate for public water supply due to a WQI value of 358.34±15.2, whereas Madalavarigudem groundwater requires treatment (WQI=59±5), and LIMAT campus groundwater indicates acceptable quality after basic treatment (WQI=20±3). Further, this study demonstrated potential in forecasting water quality parameters using a hierarchical reconciliation algorithm and predicting the WQI values using GBT, RF, and DT techniques. The predicted WQI values closely matched the obtained experimental results, further validating that Mudirajupalem groundwater is not fit for public consumption. This study opened up a scope to perform productive research on groundwater quality analysis with additional water quality parameters covering seasonal variations and employ more advanced ML techniques for predicting more accurate groundwater quality assessment for proactive water quality management.

Acknowledgments

All the individuals involved in this study are duly acknowledged with due respect. Sincere and special thanks go to the Editor and the Reviewers for providing their valuable comments and recommendations in maintaining the quality of the manuscript. Sincere and special thanks to Dr. Vita Ayu Kusuma Dewi ji for providing suggestions to improve the quality of this work.

References

[1] S. Vivek, R. Umamaheswari, P. Subashree, S. Rajakumar, P. Mukesh, P. V., S. V., L. N., G. G. Prabhu, Study on groundwater pollution and its human impact analysis using geospatial techniques in semi-urban South India, *Environmental Research*, 240, (2024), 117532. https://doi.org/10.1016/j.envres.2023.117532.

[2] Central Ground Water Board, Groundwater reports and assessments, 2008.

[3] S. S. Anchan, H. C. Shiva Prasad, Feasibility of rooftop rainwater harvesting potential—A case study of South Indian university, *Cleaner Engineering and Technology*, 4, (2021), 100206. https://doi.org/10.1016/j.clet.2021.100206.

[4] E. Murugesan, S. Shanmugamoorthy, S. Veerasamy, S. Velusamy, Groundwater hydrochemistry and its appropriateness for consumption and irrigation: Geographic and temporal variation: An integrated approach, *Urban Climate*, 49, (2023), 101482. https://doi.org/10.1016/j.uclim.2023.101482.

[5] T. Gleeson, T. Cuthbert, G. Ferguson, D. Perrone, Global groundwater sustainability, resources, and systems in the Anthropocene, *Annual Review of Earth and Planetary Sciences*, 48, (2020), 431–463. https://doi.org/10.1146/annurev-earth-071719-055251.

[6] H. Kaur, P. Garg, Urban sustainability assessment tools: A review, *Journal of Cleaner Production*, 210, (2019), 146–158. https://doi.org/10.1146/annurev-earth-071719-055251.

[7] O. F. Bulut, B. Duru, Ö. Çakmak, Ö. Günhan, F. B. Dilek, U. Yetis, Determination of groundwater threshold values: A methodological approach, *Journal of Cleaner Production*, 253, (2020), 120001. https://doi.org/10.1016/j.jclepro.2020.120001.

[8] J. Liu, Y. Peng, C. Li, Z. Gao, S. Chen, An investigation into the hydrochemistry, quality, and risk to human health of groundwater in the central region of Shandong Province, North China, *Journal of Cleaner Production*, 282, (2021), 125416. https://doi.org/10.1016/j.jclepro.2020.125416.

[9] H. Ibrahim, Z. M. Yaseen, M. Scholz, M. Ali, M. Gad, S. Elsayed, M. Khadr, H. Hussein, H. H. Ibrahim, M. Eid, A. Kovacs, S. Peter, M. Khalifa, Evaluation and prediction of groundwater quality for irrigation using integrated water quality indices, machine learning models, and GIS approaches: A representative case study, *Water*, 15, 4, (2023), 1–20. https://doi.org/10.3390/w15040694.

[10] M. G. Uddin, M. T. M. Diganta, A. M. Sajib, M. A. Hasan, M. Moniruzzaman, A. Rahman, A. I. Olbert, Assessment of hydrogeochemistry in groundwater using water quality index model and indices approaches, *Heliyon*, 9, 9, (2023), e19668. https://doi.org/10.1016/j.heliyon.2023.e19668.

[11] S. H. Mohammed, Y. G. Flores, D. A. M. Al-Manmi, V. Mikita, P. Szűcs, Assessment of groundwater level fluctuation using integrated trend analysis approaches in the Kapran sub-basin, North East Iraq, *Groundwater for Sustainable Development*, 26, (2024), 101292. https://doi.org/10.1016/j.gsd.2024.101292.

[12] Y. Fang, T. Zheng, X. Zheng, H. Peng, H. Wang, J. Xin, B. Zhang, Assessment of the hydrodynamics role for groundwater quality using integration of GIS, water quality index, and multivariate statistical techniques, *Journal of Environmental Management*, 273, (2020), 111185. https://doi.org/10.1016/j.jenvman.2020.111185.

[13] Y. Gao, H. Qian, W. Ren, H. Wang, F. Liu, F. Yang, Hydrogeochemical characterization and quality assessment of groundwater based on integrated-weight water quality index in a concentrated urban area, *Journal of Cleaner Production*, 260, (2020), 121006. https://doi.org/10.1016/j.jclepro.2020.121006.

[14] M. S. U. Hasan, A. K. Rai, Groundwater quality assessment in the Lower Ganga Basin using entropy information theory and GIS, *Journal of Cleaner Production*, 274, (2020), 123077. https://doi.org/10.1016/j.jclepro.2020.123077.

[15] J. O. Oladipo, A. S. Akinwumiju, O. S. Aboyeji, A. A. Adelodun, Comparison between fuzzy logic and water quality index methods: A case of water quality assessment in Ikare community, Southwestern Nigeria, *Environmental Challenges*, 3, (2021), 100038. https://doi.org/10.1016/j.envc.2021.100038.

[16] J. Liu, Y. Peng, C. Li, Z. Gao, S. Chen, Characterization of the hydrochemistry of water resources of the Weibei Plain, Northern China, as well as an assessment of the risk of high groundwater nitrate levels to human health, *Environmental Pollution*, 268, (2021), 115947. https://doi.org/10.1016/j.envpol.2020.115947.

[17] E. A. Hussein, C. Thron, M. Ghaziasgar, A. Bagula, M. Vaccari, Groundwater prediction using machine-learning tools, *Algorithms*, 13, 11, (2020), 300. https://doi.org/10.3390/a13110300.

[18] J. N. Apogba, G. K. Anornu, A. B. Koon, B. W. Dekongmen, E. D. Sunkari, O. F. Fynn, P. Kpiebaya, Application of machine learning techniques to predict groundwater quality in the Nabogo Basin, Northern Ghana, *Heliyon*, 10, 7, (2024), e28527. https://doi.org/10.1016/j.heliyon.2024.e28527.

[19] M. Abu, R. Musah, M. S. Zango, A combination of multivariate statistics and machine learning techniques in groundwater characterization and quality forecasting, *Geosystems and Geoenvironment*, 3, 2, (2024), 100261. https://doi.org/10.1016/j.geogeo.2024.100261.

[20] A. D. Eaton, Standard methods for the examination of water and wastewater, 21st ed., APHA–AWWA–WEF, 2005.

[21] Bureau of Indian Standards, Indian standard drinking water: Specification IS 10500-2012, 2012.

[22] R. M. Brown, N. I. McClelland, R. A. Deininger, R. G. Tozer, A water quality index: Do we dare?, *Water and Sewage Works*, 117, 10, (1972), 339–343.

[23] A. S. Eesa, Z. Orman, A. M. A. Brifcani, A novel feature-selection approach based on the cuttlefish optimization algorithm for intrusion detection systems, *Expert Systems with Applications*, 42, 5, (2015), 2670–2679. https://doi.org/10.1016/j.eswa.2014.11.009.

[24] F. J. Yang, An extended idea about decision trees, *2019 International Conference on Computational Science and Computational Intelligence*, (2019), 349–354. https://doi.org/10.1109/CSCI49370.2019.00068.

[25] J. Liang, Z. Qin, S. Xiao, L. Ou, X. Lin, Efficient and secure decision tree classification for cloud-assisted online diagnosis services, *IEEE Transactions on Dependable and Secure Computing*, 18, 4, (2021), 1632–1644. https://doi.org/10.1109/TDSC.2019.2922958.

[26] S. Kiran, G. R. Reddy, S. P. Girija, S. Venkatramulu, D. Kumar, V. Chandra Shekhar Rao, A gradient boosted decision tree with binary spotted hyena optimizer for cardiovascular disease detection and classification, *Healthcare Analytics*, 3, (2023), 100173. https://doi.org/10.1016/j.health.2023.100173.

[27] T. Hastie, R. Tibshirani, J. Friedman, The elements of statistical learning: Data mining, inference, and prediction, 2nd ed., Springer, 2009. https://doi.org/10.1007/978-0-387-84858-7.

[28] R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, 14, 2, (1995), 1137–1145.

[29] C. Bergmeir, J. M. Benítez, On the use of cross-validation for time series predictor evaluation, *Information Sciences*, 191, (2012), 192–213. https://doi.org/10.1016/j.ins.2011.12.028.

[30] R. J. Hyndman, G. Athanasopoulos, Forecasting: Principles and practice, 3rd ed., 2021.

[31] S. V. Sarath Prasanth, N. S. Magesh, K. V. Jitheshlal, N. Chandrasekar, K. Gangadhar, Evaluation of groundwater quality and its suitability for drinking and agricultural use in the coastal stretch of Alappuzha District, Kerala, India, *Applied Water Science*, 2, 3, (2012), 165–175. https://doi.org/10.1007/s13201-012-0042-5.
[32] R. E. Lamare, O. P. Singh, Application of CCME water quality index in evaluating the water quality status in limestone mining area of Meghalaya, India, *The Ecoscan*, 10, 1–2, (2016), 149–154.
[33] W.-N. Wang, A.-L. Wang, L. Chen, Y. Liu, R.-Y. Sun, Effects of pH on survival, phosphorus concentration, adenylate energy charge and Na+, K+-ATPase activities of Penaeus chinensis Osbeck juveniles, *Aquatic Toxicology*, 60, 1–2, (2002), 75–83. https://doi.org/10.1016/S0166-445X(01)00271-5.
[34] S. O. Fakayode, Impact assessment of industrial effluent on water quality of the receiving Alaro River in Ibadan, Nigeria, *Aqua International Journal of Environmental Studies*, 62, 1, (2005), 37–48.
[35] N. Khatri, S. Tyagi, D. Rawtani, M. Tharmavaram, R. D. Kamboj, Analysis and assessment of groundwater quality in Satlasana Taluka, Mehsana District, Gujarat, India through application of water quality indices, *Groundwater for Sustainable Development*, 10, (2020), 100321. https://doi.org/10.1016/j.gsd.2019.100321.
[36] W. J. Weber Jr., W. Stumm, Mechanism of hydrogen ion buffering in natural waters, *Journal of the American Water Works Association*, 55, 12, (1963), 1553–1578. https://doi.org/10.1002/j.1551-8833.1963.tb01178.x.
[37] G. Matthess, The properties of groundwater, Wiley, 1982.
[38] T. Subramani, L. Elango, S. R. Damodarasamy, Groundwater quality and its suitability for drinking and agricultural use in Chithar River Basin, Tamil Nadu, India, *Environmental Geology*, 47, 8, (2005), 1099–1110. https://doi.org/10.1007/s00254-005-1243-0.
[39] The American Well Owner, Groundwater resources and management, 2003.
[40] A. S. Murugan, C. B. Prabaharan, Fish diversity in relation to physico-chemical characteristics of Kamala Basin of Darbhanga District, Bihar, India, *International Journal of Pharmaceutical and Biological Archives*, 3, 1, (2012), 211–217.
[41] C. N. Sawyer, Chemistry for sanitary engineers, 2nd ed., McGraw-Hill, 1967.
[42] M. Ramesh, E. Dharmaraj, B. J. Raj, Physico-chemical characteristics of ground water of Manachanallur Block Trichy, Tamilnadu, India, *Advances in Applied Science Research*, 3, 3, (2012), 1709–1713.
[43] World Health Organization, Guidelines for drinking-water quality, 3rd ed., 2004.
[44] J. N. Marfo, J. A. Quaye-Ballard, S. O. Kwakye, K. Obeng, A. Arko-Adjei, N. L. Quaye-Ballard, R. N. A. Quao, Groundwater quality and potential analysis using geospatial techniques: The case of Ashanti Region in Ghana, *Heliyon*, 10, 6, (2024), e27545. https://doi.org/10.1016/j.heliyon.2024.e27545.
[45] P. R. Kumar, S. S. Gowd, C. Krupavathi, Groundwater quality evaluation using water quality index and geospatial techniques in parts of Anantapur District, Andhra Pradesh, South India, *HydroResearch*, 7, (2024), 86–98. https://doi.org/10.1016/j.hydres.2024.01.001.
[46] C. Ramakrishna, D. M. Rao, K. S. Rao, N. Srinivas, Studies on groundwater quality in slums of Visakhapatnam, Andhra Pradesh, *Asian Journal of Chemistry*, 21, 6, (2009), 4246–4250.
[47] B. Vamsi, B. Chandra Sekhar, V. Vakdevi, M. Rohit, M. Ibrahim, P. Uday Kiran, Ch. Srinivas, M. Gopi Kiran, Groundwater quality analysis and water quality index prediction by means of machine learning methods, *SPC Journal of Environmental Sciences*, 7, 1, (2025), 1–8. https://doi.org/10.14419/wkcxnn03.
[48] X. Wang, F. Zhang, J. Ding, Evaluation of water quality based on a machine learning algorithm and water quality index for the Ebinur Lake Watershed, China, *Scientific Reports*, 7, 1, (2017), 12853. https://doi.org/10.1038/s41598-017-12853-y.
[49] B. F. Sami, S. D. Latif, A. N. Ahmed, M. F. Chow, M. A. Murti, A. Suhendi, B. H. Ziyad Sami, J. K. Wong, A. H. Birima, A. El-Shafie, Machine learning algorithm as a sustainable tool for dissolved oxygen prediction: A case study of Feitsui Reservoir, Taiwan, *Scientific Reports*, 12, 1, (2022), 6969. https://doi.org/10.1038/s41598-022-06969-z.
[50] J. L. Ubah, L. C. Orakwe, K. N. Ogbu, J. L. Awu, I. E. Ahaneku, E. C. Chukwuma, Forecasting water quality parameters using artificial neural networks for irrigation purposes, *Scientific Reports*, 11, 1, (2021), 24438. https://doi.org/10.1038/s41598-021-04062-5.