Environment and Ecology 44 (2) 385—400, April—June 2026
Article DOI: https://doi.org/10.60151/envec/EHDF3490
ISSN: 0970-0420 (Print), ISSN: 3049-3625 (Online)

Prediction of WQI Using an Artificial Neural Network Model for Drinking Water Management in the Therlam Region, Vizianagaram District, India

Rupakumari Gedala, Srinivasa Rao. G.V.R.

Received 2 March 2026, Accepted 22 May 2026, Published on 17 June 2026

ABSTRACT

Groundwater quality monitoring in hard-rock aquifer systems is essential for sustainable resource management in rural regions dependent on groundwater for drinking purposes. The present study evaluates groundwater quality using the weighted arithmetic Water Quality Index (WQI) method and develops an Artificial Neural Network (ANN)-based framework for rapid WQI estimation in the Therlam region of Vizianagaram, Andhra Pradesh. A total of 828 groundwater samples were collected from 23 monitoring stations during November 2018–October 2021 and analyzed for fifteen physicochemical parameters following American Public Health Association (2017) standard procedures. WQI classification indicated that 30.43% of samples belong to the good to excellent category, 47.83% to poor, 17.39% to very poor, and 4.35% to unsuitable category, indicating moderate hydrochemical variability within the study area. To estimate WQI efficiently from routinely monitored hydrochemical parameters, a feed-forward ANN model with a 15–5–1 architecture trained using the Levenberg–Marquardt algorithm was developed. Approximately 70%, 15%, and 15% of the dataset were used for training, validation, and testing, respectively. Multicollinearity among input variables was evaluated using Variance Inflation Factor (VIF) analysis prior to model implementation. Model performance was further compared with Multiple Linear Regression (MLR) and Random Forest (RF) approaches. The ANN model achieved satisfactory predictive performance (R^2 = 0.9658; RMSE = 0.4964, MAE = 0.3624), outperforming the baseline models. The developed ANN framework is intended as a rapid computational tool for WQI estimation rather than an independent forecasting model. The study highlights the applicability of machine-learning-assisted groundwater quality assessment in hard-rock aquifer regions while emphasizing the need for inclusion of hydro-meteorological and anthropogenic variables in future predictive studies. However, the developed ANN framework represents a site-specific surrogate estimation model and should not be interpreted as an independent groundwater-quality forecasting system.

Keywords Groundwater quality, Water quality Index, Artificial neural network, Hard-rock aquifer, Hydrochemistry.

Rupakumari Gedala[1]*, Srinivasa Rao. G.V.R.[2]

[1]PhD Research Scholar, [2]Professor

[1,2]Department of Civil Engineering, AU College of Engineering (A), Andhra University, Visakhapatnam 530003, Andhra Pradesh India

Email: g.rupakumari@gmail.com
*Corresponding author

INTRODUCTION

Groundwater is the principal source of drinking water in rural regions of India, particularly within hard-rock aquifer terrains where surface-water availability is limited and seasonal. In such hydrogeological environments, groundwater chemistry is controlled by lithology, mineral–water interaction, recharge processes, residence time, and anthropogenic activities including agriculture and domestic waste disposal. (Gao *et al.* 2020, Kumar *et al.* 2021, Subba Rao *et al.* 2020, Wu *et al.* 2019). These factors frequently produce considerable spatial and temporal variability in groundwater quality, thereby influencing its suitability for drinking and domestic purposes. Consequently, systematic assessment and monitoring of groundwater quality are essential for sustainable water-resource management and protection of public health.

The Water Quality Index (WQI) is widely used for evaluating overall groundwater quality by integrating multiple physicochemical parameters into a single numerical value. Among various methods, the weighted arithmetic WQI approach proposed by Horton (1965) and later modified by Brown (1972) has been extensively applied because of its simplicity, interpretability, and suitability for drinking-water assessment. The method enables classification of groundwater into distinct quality categories ranging from excellent to unsuitable, thereby facilitating interpretation of hydrochemical data for water-resource managers and policymakers.

Several investigations in India have successfully employed WQI techniques to assess groundwater quality under different hydrogeological conditions, including studies from Karnataka (Ramakrishnaiah *et al.* 2009), Nagpur (Rajankar *et al.* 2009), Tamil Nadu (Vasanthavigar *et al.* 2010), Greater Visakhapatnam (Swarnalatha and Nageswara Rao 2010, Srinivasa Rao and Nageswara Rao 2013), and parts of Vizianagaram district (Pavan Kumar *et al.* 2017). Review-based assessments have also summarized groundwater-quality trends and contamination issues across India (Renu Nayar 2020, Basavaraja *et al*. 2025). Although these studies provide valuable information regarding groundwater suitability, conventional WQI methods are primarily descriptive in nature.

Recent studies have explored the application of data-driven approaches such as Artificial Neural Networks (ANN), Random Forest (RF), and Support Vector Machine (SVM) models for groundwater-quality assessment and WQI estimation. Among these approaches, ANN models have received considerable attention because of their ability to approximate nonlinear relationships among hydrochemical variables (Maier *et al,* 2000). However, recent reviews have also highlighted several methodological limitations in groundwater-quality modeling studies, including overfitting, data leakage, insufficient validation procedures, and lack of independent testing datasets (Muthukumaran and Vinoth Kumar 2022, Ghadai *et al.* 2022, Haggerty *et al.* 2023, Ahn *et al.* 2023).

Despite increasing application of machine-learning techniques, relatively few investigations have utilized long-term hydrochemical monitoring datasets from inland hard-rock aquifer systems of semi-arid rural regions in India. Moreover, many previous ANN-based studies relied on limited datasets and simplified validation strategies. The present study therefore integrates multi-year groundwater-quality monitoring with ANN-based WQI estimation in the Therlam region of Vizianagaram, Andhra Pradesh.

Unlike conventional forecasting studies, the ANN framework developed in the present investigation is intended as a computational surrogate model for rapid WQI estimation using routinely monitored hydrochemical parameters rather than as an independent groundwater-quality forecasting system. To improve methodological reliability, the study incorporates structured dataset partitioning, validation monitoring, and comparative assessment with baseline statistical and machine-learning approaches.

The specific objectives of the present study are to:

1 Evaluate groundwater quality using the weighted arithmetic WQI method.
2 Develop and validate an ANN-based framework for rapid WQI estimation.
3 Compare ANN performance with baseline statis tical and machine-learning models.
4 Assess groundwater-quality conditions within the hard-rock aquifer system of the study area.

MATERIALS AND METHODS

Study area

The study area (Fig.1) comprises Therlam, Merakamudidam, Garividi, and Cheepurupalli mandals of Vizianagaram district, Andhra Pradesh, covering an extent of area approximately 597 km² between 18°12′00″–18°33′36″ N latitude and 83°24′36″–83°42′00″ E longitude. The region forms part of the Eastern Ghats Mobile Belt and is predominantly underlain by hard-rock aquifer systems associated with the Khondalite Group formations of the Eastern Ghats terrain (K. Krishna Kumar 2012). Groundwater occurrence is mainly confined to weathered and fractured crystalline formations.

The climate of the region is tropical monsoonal, characterized by distinct wet and dry seasons, with groundwater recharge primarily controlled by southwest monsoon precipitation. Seasonal recharge variations, groundwater residence time, and rock–water interactions significantly influence groundwater chemistry within the aquifer system. Groundwater constitutes the principal source of drinking water for rural and tribal communities inhabiting the study area. The location map was prepared using QGIS software.

Sampling and laboratory analysis

A total of 828 groundwater samples were collected from 23 monitoring stations during a continuous

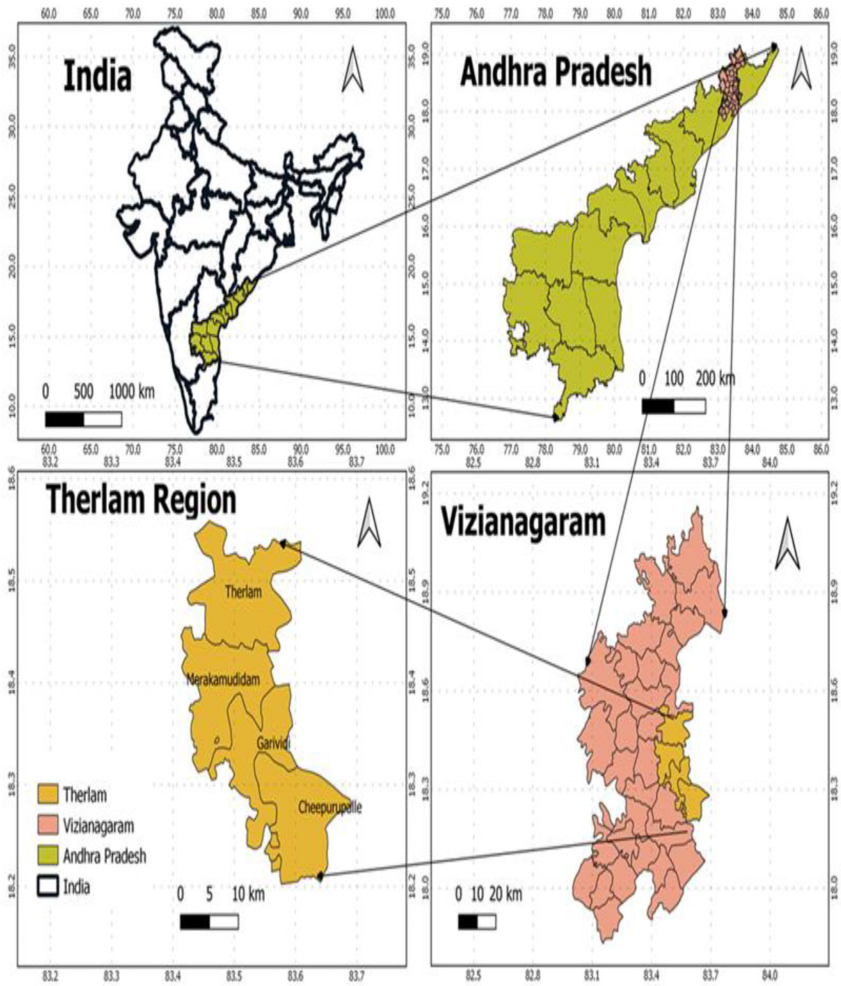

Fig. 1. Study area map.

36-month period extending from November 2018 to October 2021. Groundwater samples were collected monthly from each monitoring location to capture spatial and temporal hydrochemical variability. The geographical coordinates and sampling details of all monitoring stations are presented in Table 1.

Groundwater samples were collected in pre-cleaned polyethylene bottles following standard sampling protocols. The bottles were washed with non-ionic detergent, rinsed with tap water, treated with 1:1 hydrochloric acid solution, and finally rinsed thoroughly using deionized water. Prior to sample collection, each bottle was rinsed three times with the respective groundwater sample to minimize contamination. Tube wells were pumped continuously for approximately five minutes before sampling to obtain representative aquifer water.

Physicochemical analyses were carried out following the standard procedures prescribed by American Public Health Association (2017). pH and electrical conductivity (EC) were measured using a digital pH meter (Elico LI-120) and conductivity meter (Elico CL-351), respectively. Total dissolved solids (TDS) were determined by the gravimetric method. Total hardness (TH), total alkalinity (TA), calcium (Ca^{2+}), magnesium (Mg^{2+}), chloride (Cl^{-}), carbonate (CO_3^{2-}), and bicarbonate (HCO_3^{-}) were analyzed using standard titrimetric methods.

Nitrate (NO_3^{-}) concentration was determined by the phenol disulphonic acid method using a UV–Visible spectrophotometer (Elico SL-177) with a 1-cm quartz cell. Fluoride (F^{-}) was analyzed using the SPADNS method, whereas sulphate (SO_4^{2-}) concentration was determined by the turbidimetric method employing standard barium chloride solution. Sodium (Na^{+}) and potassium (K^{+}) concentrations were measured using a flame photometer (Elico CL-361).

Fifteen physicochemical parameters, namely pH, EC, TDS, TH, TA, Ca^{2+}, Mg^{2+}, Na^{+}, K^{+}, NO_3^{-}, Cl^{-}, F^{-}, SO_4^{2-}, CO_3^{2-}, and HCO_3^{-}, were used for Water Quality Index (WQI) computation and ANN model development.

Calculation of water quality index (WQI)

The Weighted Arithmetic Water Quality Index

Table 1. Groundwater sampling locations and coordinates.

Sl. No.	Sample Id	Sampling Station	Mandal	Latitude (N)	Longitude(E)
1	S1	Amity	Therlam	18°29’52”	83°31’50”
2	S2	Gangannapadu	Therlam	18°29’41”	83°28’47”
3	S3	Kusumuru	Therlam	18°32’38”	83°29’40”
4	S4	Nandabalaga	Therlam	18°32’42”	83°28’37”
5	S5	Nandigam	Therlam	18°30’54”	83°29’30”
6	S6	Therlam	Therlam	18°28’33”	83°30’25”
7	S7	Billalavalasa	Merakamudidam	18°21’58”	83°30’03”
8	S8	Garugubilli	Merakamudidam	18°23’19”	83°33’20”
9	S9	Ippalavalasa	Merakamudidam	18°20’27”	83°29’50”
10	S10	Uttaravalli	Merakamudidam	18°27’42”	83°28’38”
11	S11	Yadika	Merakamudidam	18°22’10”	83°33’40”
12	S12	Devada	Garividi	18°15’20”	83°33’33”
13	S13	Yenuguvalasa	Garividi	18°19’16”	83033’06”
14	S14	Garividi	Garividi	18°16’49”	83°31’57”
15	S15	Neeladripuram	Garividi	18°20’14”	83°31’09”
16	S16	Seripeta	Garividi	18°15’32”	83°31’13”
17	S17	Alajangi	Cheepurupalli	18°18’00”	83°36’35”
18	S18	Devarapodilam	Cheepurupalli	19°00’42”	83°31’19”
19	S19	Karlam	Cheepurupalli	18°17’24”	83°38’43”
20	S20	Nimmavalasa	Cheepurupalli	18°45’30”	83°50’33”
21	S21	Parla	Cheepurupalli	18°14’54”	83°34’57”
22	S22	Peripi	Cheepurupalli	18°15’56”	83°36’47”
23	S23	Ramalingpuram	Cheepurupalli	18°18’44”	83°36’26”

Table 2. Water Quality rating as per WA-WQI model Brown *et al.* (1972).

WQI Value	Rating of water quality	Suitability for drinking
0-25	Excellent	Fit for drinking
26-50	Good	Acceptable for drinking
51-75	Poor	Moderately polluted
76-100	Very Poor	Excessively polluted
Above100	Unsuitable for drinking purpose	Unfit for drinking

(WAWQI) method was adopted to evaluate groundwater suitability for drinking purposes because of its simplicity and ability to integrate multiple water-quality variables into a single representative value. Standard permissible limits recommended by the World Health Organization and Bureau of Indian Standards were used for WQI computation. Classification of Water quality ranking prescribed by Brown *et al*. 1972 was presented in the Table 2.

The overall WQI was calculated using:

$$WQI = \frac{\sum Q_i W_i}{\sum W_i} \qquad \text{(Eq. 1)}$$

The quality rating scale (Qi) for each parameter is calculated by using this expression:

$$Qi = 100 \left(\frac{V_i - V_o}{S_i - V_o}\right) \qquad \text{(Eq. 2)}$$

Where, V_i= estimated concentration of i^{th} parameter in the analyzed water, V_o=The ideal value of this parameter in pure water, V_o= 0 (except pH =7.0), S_i=Standard value of i^{th} parameter The unit weight (W_i) for each water quality parameter is calculated by using the following formula:

$$W_i = \frac{K}{S_i} \qquad \text{(Sq. 3)}$$

Where, K= proportionality constant and can also be calculated by using the following equation:

$$K = \frac{1}{\sum \left(\frac{1}{si}\right)} \qquad \text{(Sq. 4)}$$

Artificial neural net work (ANN) model development

A feed-forward Multilayer Perceptron (MLP) Artificial Neural Network (ANN) was developed to estimate WQI values using routinely monitored hydrochemical parameters as predictor variables. The ANN framework was designed as a computational surrogate model for rapid WQI estimation rather than as an independent forecasting model incorporating external environmental variables.

The complete dataset consisted of 828 groundwater samples collected from 23 monitoring stations during the 36-month monitoring period. To improve methodological reliability and minimize information leakage, the dataset was partitioned chronologically. Data corresponding to the first 30 months (n = 690) were used for model development, whereas data from the final six months (n = 138) were retained exclusively for independent testing under unseen temporal conditions.

The calibration dataset was further randomly partitioned within MATLAB into training (70%), validation (15%), and internal testing (15%) subsets using the default random division function. The validation subset was used for early stopping and monitoring of model generalization during training.

The optimized ANN architecture consisted of 15 input neurons, one hidden layer containing five neurons, and one output neuron (15–5–1 configuration) (Fig. 2). A tangent sigmoid (tansig) activation function was employed in the hidden layer, whereas a linear (purelin) activation function was applied in the output layer.

Prior to model training, all input variables were normalized between 0 and 1 using min–max scaling to improve computational stability and convergence efficiency.

The optimal hidden-layer configuration was selected through iterative experimentation by evaluating network structures containing 3–10 hidden neurons and comparing validation errors. The five-neuron hidden layer provided the best compromise between predictive performance and generalization capability, while larger architectures showed negligible improvement and increased susceptibility to overfitting.

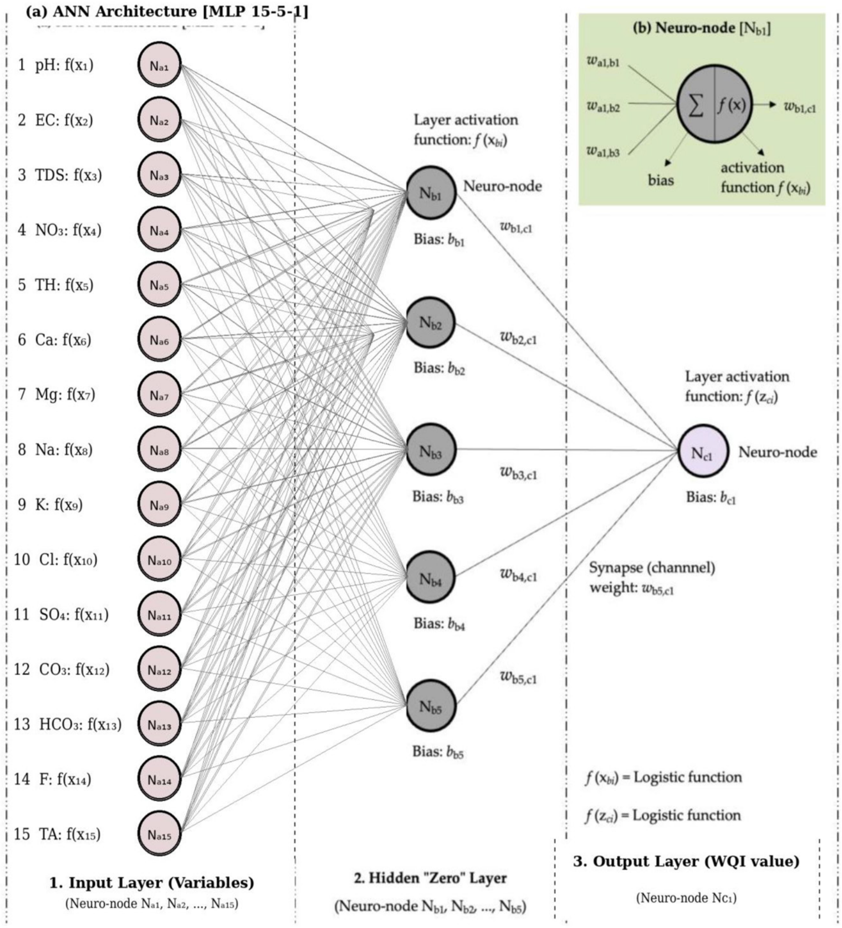

Fig. 2. Architecture of the feed-forward ANN model for WQI prediction.

Model training was performed using the Levenberg–Marquardt backpropagation algorithm implemented in MATLAB Neural Network Toolbox (Model 2018a, MathWorks, USA). The final ANN configuration contained 86 adjustable parameters, which remained substantially lower than the effective training sample size, thereby reducing the possibility of over-parameterization.

For comparative assessment, Multiple Linear Regression (MLR) and Random Forest (RF) models were additionally implemented.

Performance evaluation metrics

Model performance was evaluated using coefficient of determination (R^2), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Bias Error (MBE).

The statistical expressions are given below:
Coefficient of Determination

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (O_i - \hat{P}_i)^2}{\sum_{i=1}^{n} (o_i - \bar{o})^2} \quad \text{(Rq. 5)}$$

Root Mean Square Error

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (O_i - \hat{P}_i)^2} \quad \text{(Rq. 6)}$$

Mean Absolute Error

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left| O_i - \hat{P}_i \right| \qquad \text{(Rq. 7)}$$

Mean Bias Error

$$MBE = \frac{1}{n}\sum_{i=1}^{n} (Pi - Oi) \qquad \text{(Rq. 8)}$$

Where Oi = observed WQI; Pi = predicted WQI; Ō = mean observed WQI; n = number of samples

Multicollinearity assessment

Prior to ANN implementation, multicollinearity among hydrochemical variables was evaluated using Variance Inflation Factor (VIF) analysis to identify redundancy among predictor variables and improve model robustness. The VIF for each predictor variable was calculated as:

$$VIF_i = 1/1 - R_i^2$$

Where: R_i^2 = coefficient of determination obtained by regressing the i^{th} predictor against all remaining predictor variables.

Variables exhibiting VIF values less than 10 were considered acceptable for model implementation. Moderate inter-parameter correlation was expected because hydrochemical variables in groundwater systems are commonly influenced by similar geochemical processes such as mineral dissolution, ion exchange, and evaporation–concentration mechanisms.

RESULTS AND DISCUSSION

Physicochemical characteristics of groundwater

The analytical results of groundwater samples collected from twenty-three monitoring stations (S1–S23) are summarized in Table 3. The groundwater of the study area is generally alkaline in nature, with pH values ranging from 7.05 to 8.03, indicating slightly alkaline hydrochemical conditions within permissible drinking-water limits prescribed by the World Health Organization (2022) /Bureau of Indian Standards BIS-10500 2012.

Electrical conductivity (EC) values varied considerably from 594 to 4124 µS/cm, indicating substantial spatial heterogeneity in dissolved ionic concentration. Elevated EC values observed at sampling stations S1, S4, and S5 suggest comparatively higher mineralization and salinity conditions. Such variations are commonly associated with prolonged water–rock interaction, evapoconcentration processes, and localized anthropogenic influences.

Total dissolved solids (TDS) concentrations ranged between 400 and 2658.6 mg/L, with an average concentration of approximately 1320 mg/L. According to conventional groundwater classification, waters containing TDS values exceeding 1000 mg/L are considered brackish. Several sampling stations exceeded this threshold, indicating localized salinity problems that may adversely affect palatability and domestic usability.

Total hardness (TH) values ranged from 130 to 663 mg/L, with a mean concentration of approximately 392 mg/L, indicating predominantly hard to very hard groundwater. Elevated hardness is primarily attributable to dissolution of calcium- and magnesium-bearing minerals within the hard-rock aquifer system. Total alkalinity (TA) varied from 100 to 571 mg/L and was strongly associated with bicarbonate (HCO_3^-) dominance (100–566 mg/L), suggesting that carbonate weathering and mineral dissolution processes exert major control on groundwater chemistry.

Major ions and salinity indicators

Sodium concentrations varied between 47 and 483 mg/L, with comparatively higher values recorded at stations S1, S5, S18, and S19. Elevated sodium concentrations, together with increased chloride levels (72–731.8 mg/L), indicate localized salinity enrichment possibly associated with irrigation return flow, evaporative concentration, or anthropogenic contamination sources. Nevertheless, chloride concentrations remained within the permissible drinking-water limit of 1000 mg/L at all sampling locations.

Sulphate concentrations were relatively low, ranging from 1 to 93.6 mg/L, and remained well below the permissible limit of 400 mg/L, indicating

Table 3. Characteristics of Physico-Chemical Parameters of Groundwater Sample.

Sample ID	S1		S2		S3		S4		S5	
Parameter	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
pH	7.05	7.63	7.05	7.73	7.05	7.73	7.05	7.53	6.95	7.63
EC	3788.00	3851.00	373.00	3476.00	2466.00	2538.00	3275.00	3335.00	4069.00	4124.00
TDS	2444.00	2483.60	2194.00	2243.60	1598.00	1643.60	2116.00	2153.60	2624.00	2658.60
NO3	30.00	35.60	24.00	30.60	30.00	35.60	10.00	16.60	54.00	59.60
TH	535.00	563.00	527.00	553.00	405.00	438.00	495.00	523.00	640.00	663.00
Ca	101.00	106.60	172.00	177.60	100.00	105.60	100.00	106.60	99.00	105.60
Mg	43.00	50.60	20.00	25.60	38.00	43.60	56.00	61.60	69.00	74.60
Na	473.00	483.20	398.00	407.20	108.00	121.20	208.00	217.20	466.00	475.20
K	94.90	96.06	79.80	80.76	20.90	23.06	42.40	45.06	90.90	95.06
Cl	717.00	731.80	605.00	621.80	162.00	186.80	317.00	336.80	697.00	716.80
SO4	54.00	59.60	58.00	63.60	46.00	51.60	88.00	93.60	50.00	55.60
CO3	12.00	18.60	125.00	140.60	9.00	15.60	10.00	15.60	99.00	112.60
HCO3	420.00	486.00	250.00	296.00	400.00	446.00	420.00	456.00	360.00	406.00
F	1.06	1.23	0.52	0.66	0.29	0.47	0.60	0.75	0.52	0.65
TA	440.00	486.00	260.00	316.00	420.00	466.00	440.00	476.00	370.00	426.00

Sample ID	S6		S7		S8		S9		S10	
Parameter	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
pH	7.05	7.63	7.03	7.63	7.35	7.93	6.97	7.53	7.15	7.53
EC	1506.00	1585.00	1392.00	1466.00	1852.00	1930.00	594.00	668.00	1710.00	1779.00
TDS	984.00	1033.60	911.00	957.60	1205.00	1254.60	400.00	446.60	1114.00	1157.60
NO3	14.00	20.60	0.00	6.60	10.50	16.60	8.00	14.60	20.00	26.60
TH	267.00	293.00	400.00	428.00	505.00	538.00	130.00	158.00	430.00	458.00
Ca	102.00	107.60	141.00	146.60	46.00	51.60	34.00	40.60	62.00	67.60
Mg	63.00	69.60	61.00	69.60	93.00	98.60	11.00	16.60	64.00	69.60
Na	69.00	80.20	130.00	141.20	68.00	77.20	47.00	59.20	63.00	76.20
K	14.00	15.56	25.10	28.26	12.90	14.06	9.90	11.26	12.90	14.06
Cl	107.00	121.80	202.00	221.80	102.00	121.80	72.00	91.80	97.00	117.80
SO4	70.00	75.60	1.00	9.60	17.10	23.60	1.00	8.60	17.00	22.60
CO3	91.00	100.60	149.00	190.60	64.00	80.60	19.00	30.60	34.00	45.60
HCO3	310.00	346.00	340.00	396.00	400.00	436.00	100.00	131.00	375.00	406.00
F	0.57	0.73	0.27	0.39	0.58	0.71	0.01	0.11	0.44	0.58
TA	320.00	376.00	360.00	416.00	420.00	456.00	100.00	151.00	390.00	426.00

Sample ID	S11		S12		S13		S14		S15	
Parameter	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
pH	7.45	8.03	7.15	7.73	7.03	7.48	7.25	7.83	7.21	7.83
EC	1783.00	1861.00	1663.00	1741.00	1622.00	1694.00	1136.00	1215.00	910.00	963.00
TDS	1161.00	1210.60	1089.00	1133.60	1058.00	1103.60	747.00	796.60	602.00	635.60

Table 3. Continued.

Sample ID	S11		S12		S13		S14		S15	
Parameter	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
NO3	18.00	24.80	17.00	22.60	15.00	20.60	6.00	11.60	17.50	23.60
TH	460.00	488.00	480.00	508.00	385.00	408.00	370.00	398.00	195.00	223.00
Ca	60.00	65.60	59.00	64.60	60.00	65.60	49.00	54.60	35.00	40.60
Mg	75.00	80.60	80.00	85.60	55.00	60.60	87.00	92.60	24.00	29.60
Na	67.00	77.20	138.00	148.20	103.00	116.20	89.00	99.20	52.00	67.20
K	12.90	14.96	17.10	18.66	19.50	21.26	18.00	19.86	10.90	13.46
Cl	102.00	121.80	212.00	231.80	157.00	181.80	137.00	156.80	82.00	106.80
SO4	13.00	19.10	17.00	22.60	48.00	53.60	70.00	75.60	10.00	15.60
CO3	50.00	60.60	29.00	40.60	17.00	22.60	39.00	50.60	9.00	12.60
HCO3	400.00	436.00	430.00	456.00	325.00	351.00	370.00	396.00	180.00	206.00
F	0.64	0.80	0.60	0.73	0.61	0.73	0.74	0.88	0.24	0.35
TA	400.00	451.00	440.00	481.00	335.00	371.00	69.00	416.00	190.00	231.00

Sample ID	S16		S17		S18		S19		S20	
Parameter	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
pH	7.03	7.59	7.03	7.53	7.45	8.03	7.35	8.03	7.31	7.83
EC	1420.00	1498.00	1663.00	1747.00	1585.00	1655.00	2006.00	2086.00	1647.00	1735.00
TDS	929.00	978.60	1084.00	1147.60	1034.00	1078.60	1304.00	1354.60	1074.00	1129.60
NO3	19.00	25.60	23.00	29.60	34.00	39.60	33.00	39.60	28.00	33.60
TH	380.00	408.00	395.00	428.00	305.00	328.00	590.00	618.00	300.00	328.00
Ca	28.00	33.60	171.00	176.60	25.20	31.60	213.00	218.60	92.00	97.60
Mg	75.00	80.60	76.00	81.60	82.00	88.10	89.00	96.60	13.00	18.60
Na	66.00	77.20	183.00	194.20	213.00	224.20	193.00	204.20	141.00	152.20
K	13.50	15.46	36.90	39.56	42.70	46.46	37.30	40.96	27.70	31.06
Cl	102.00	121.80	282.00	301.80	327.00	346.80	297.00	316.80	217.00	236.80
SO4	10.00	15.60	40.00	45.60	68.00	73.60	9.00	14.60	14.00	19.60
CO3	8.00	12.60	189.00	204.60	20.00	28.60	219.00	232.60	19.00	26.60
HCO3	400.00	426.00	390.00	436.00	330.00	366.00	520.00	566.00	265.00	296.00
F	0.59	0.77	0.27	0.38	0.84	0.98	0.46	0.60	0.28	0.44
TA	400.00	446.00	390.00	426.00	332.00	376.00	535.00	571.00	275.00	311.00

Sample ID	S21		S22		S23		Permissible values (Si)
Parameter	Min	Max	Min	Max	Min	Max	
pH	7.31	7.83	7.15	7.53	7.05	7.53	8.5
EC	1545.00	1608.00	1756.00	1826.00	1366.00	1432.00	1000
TDS	1009.00	1048.60	1144.00	1187.60	894.00	935.60	600
NO3	23.00	30.10	22.00	27.60	25.00	31.60	50
TH	380.00	408.00	350.00	378.00	315.00	343.00	500
Ca	60.00	65.60	62.00	67.60	25.20	30.60	200

Table 3. Continued.

Sample ID	S21		S22		S23		Permissible values (Si)
Parameter	Min	Max	Min	Max	Min	Max	
Mg	55.00	60.60	45.00	50.60	82.00	88.10	150
Na	67.00	77.20	68.00	76.20	60.00	71.20	200
K	14.10	15.96	14.00	15.96	12.60	14.46	20
Cl	102.00	121.80	102.00	116.80	92.00	111.80	250
SO4	19.00	25.60	16.50	22.60	68.00	73.60	250
CO3	41.00	44.60	29.00	34.60	19.00	24.60	30
HCO3	330.00	351.00	315.00	341.00	340.00	366.00	500
F	0.58	0.69	0.36	0.51	0.58	0.72	1.2
TA	335.00	371.00	310.00	361.00	350.00	386.00	200

All parameters are expressed in mg/L except pH (no units) and EC=µS/cm.

minimal influence of sulphate-bearing lithology within the study area.

Nitrate and Fluoride Assessment

Nitrate concentrations ranged from 0 to 59.6 mg/L, with a mean concentration of approximately 23 mg/L. Elevated nitrate concentrations observed at certain locations may be associated with agricultural runoff, fertilizer application, septic-system leakage, and domestic wastewater infiltration. In particular, nitrate concentration at station S5 exceeded the permissible limit of 45–50 mg/L, indicating potential anthropogenic contamination and possible public-health concerns such as methemoglobinemia.

Fluoride concentrations varied from 0.01 to 0.98 mg/L and remained within the desirable drinking-water limit of 1.0 mg/L across all sampling stations. Consequently, groundwater from the study area does not presently pose significant fluorosis-related health risks.

Spatial variability and overall, groundwater quality

Considerable spatial variability was observed among hydrochemical parameters, particularly EC, TDS, TH, sodium, and chloride concentrations. Sampling stations S1, S4, and S5 represent comparatively mineralized zones characterized by elevated salinity and hardness, whereas stations S9, S15, and S23 exhibited relatively better groundwater quality.

Overall, groundwater quality within the study area is primarily controlled by geogenic processes including mineral dissolution, silicate weathering, and carbonate weathering, while localized anthropogenic impacts are reflected by nitrate enrichment at select locations.

Although most hydrochemical parameters remain within national and international permissible limits, site-specific treatment methods such as reverse osmosis, softening, or desalination are recommended in highly mineralized zones to improve potable suitability.

Water quality index (WQI) Assessment

The Water Quality Index (WQI) was calculated for all 828 groundwater samples collected from 23 monitoring stations during the period from November 2018 to October 2021. The computed WQI values exhibited substantial spatial and temporal variability, reflecting differences in hydrochemical characteristics across the study area.

The average WQI values ranged from 11.24 to 104.44 (Table 4). The highest WQI value (104.44) was recorded at station S1, categorizing the groundwater as unsuitable for drinking purposes, whereas

Table 4. Water Quality Index (WQI) Values and Groundwater Quality Classification of Sampling Stations.

Sample ID	Yearly Average WQI Values			Average WQI values	WQI ranking
	Year 1	Year 2	Year3		
S1	103.23	105.16	104.92	104.44	Unsuitable
S2	75.38	77.30	77.06	76.58	Very Poor
S3	36.07	38.00	37.76	37.28	Good
S4	59.14	61.07	60.83	60.35	Poor
S5	77.68	79.61	79.37	78.88	Very Poor
S6	58.91	60.84	60.60	60.12	Poor
S7	48.30	50.22	49.98	49.50	Good
S8	57.89	59.82	59.58	59.09	Poor
S9	10.04	11.97	11.72	11.24	Excellent
S10	44.36	46.29	46.05	45.56	Good
S11	62.20	64.13	63.89	63.40	Poor
S12	55.40	57.33	57.09	56.61	Poor
S13	52.48	54.41	54.16	53.68	Poor
S14	65.68	67.61	67.37	66.89	Poor
S15	26.69	28.62	28.38	27.90	Good
S16	50.96	52.89	52.65	52.17	Poor
S17	54.25	56.18	55.94	55.46	Poor
S18	78.20	80.13	79.89	79.41	Very Poor
S19	74.61	76.54	76.30	75.82	Very Poor
S20	37.20	39.13	38.89	38.41	Good
S21	54.36	56.29	56.04	55.56	Poor
S22	38.12	40.04	39.80	39.32	Good
S23	50.02	51.95	51.71	51.23	Poor

the lowest WQI value (11.24) was observed at station S9, corresponding to the excellent water-quality category. Such variability indicates localized hydrochemical deterioration within certain parts of the aquifer system.

Temporal Variation in WQI

Year-wise analysis of WQI values indicated a gradual increasing trend from Year 1 to Year 3 across most sampling stations. Although the magnitude of change was moderate, the consistent increase suggests progressive deterioration in groundwater quality during the monitoring period.

This trend may be associated with: expansion of agricultural activities, increasing groundwater exploitation, return flow from irrigated fields, domestic wastewater infiltration, and Inadequate waste-management practices.

Seasonal recharge fluctuations and evapoconcentration effects may also contribute to temporal variability in dissolved ionic concentrations and WQI values.

Spatial Distribution of WQI Categories

Significant spatial variation in groundwater suitability was observed throughout the study area. Stations S1, S2, S5, S18, and S19 exhibited comparatively higher WQI values corresponding to very poor and unsuitable categories, indicating elevated hydrochemical contamination. Conversely, stations S3, S9, S15, S20, and S22 showed relatively better groundwater quality conditions.

Overall WQI classification revealed that: 30.43% of groundwater samples fall under excellent and good categories, 47.83% fall under the poor category, 17.39% fall under the very poor category, and 4.35% fall under the unsuitable category.

The predominance of poor-quality groundwater suggests moderate deterioration in groundwater suitability across the study region. Approximately half of the groundwater sources require some degree

Table 5. Comparison Between Experimented and ANN Predicted WQI Values during.

			Testing					
Sl No.	Sample ID	Month & Year	May-21	Jun-21	Jul-21	Aug-21	Sep-21	Oct-21
		Experimented	110.27	108.93	102.15	100.69	100.60	101.88
1	S1	Ann Predicted	110.05	109.41	102.71	101.32	100.11	103.14
		Experimented	81.03	79.97	74.72	75.65	73.24	74.96
2	S2	Ann Predicted	80.62	80.05	74.62	75.72	73.27	75.02
		Experimented	40.92	44.33	36.10	34.02	32.51	32.76
3	S3	Ann Predicted	39.93	43.31	39.42	34.03	32.56	32.67
		Experimented	64.63	64.85	59.03	57.49	56.26	58.19
4	S4	Ann Predicted	64.95	65.01	59.10	57.53	54.68	58.25
		Experimented	83.13	83.58	76.68	77.62	76.46	76.12
5	S5	Ann Predicted	82.47	83.77	76.57	77.60	76.07	75.93
		Experimented	64.08	65.98	57.17	56.17	55.75	57.34
6	S6	Ann Predicted	65.18	65.20	57.24	55.71	54.67	57.19
		Experimented	53.13	54.63	48.98	48.83	47.53	48.23
7	S7	Ann Predicted	53.24	54.85	48.88	48.67	47.63	48.04
		Experimented	60.88	62.21	57.74	56.33	55.92	57.12
8	S8	Ann Predicted	60.91	62.29	57.86	56.14	55.46	56.99
		Experimented	15.20	12.29	9.57	8.98	10.09	10.87
9	S9	Ann Predicted	14.57	12.12	9.60	9.08	10.17	10.79
		Experimented	49.50	50.46	43.18	42.61	41.81	44.27
10	S10	Ann Predicted	49.14	50.69	43.06	42.54	40.19	44.03
		Experimented	66.90	68.69	59.74	59.18	59.78	60.05
11	S11	Ann Predicted	66.57	68.77	59.65	58.84	59.70	59.99
		Experimented	59.56	60.50	54.28	53.62	55.06	54.97
12	S12	Ann Predicted	57.07	60.74	53.79	53.45	53.76	54.88
		Experimented	57.89	57.18	52.71	52.54	51.33	52.39
13	S13	Ann Predicted	57.12	57.33	52.87	52.89	50.97	52.63
		Experimented	70.01	70.95	64.83	65.12	64.13	64.72
14	S14	Ann Predicted	70.23	70.80	64.91	65.39	65.27	65.20
		Experimented	29.44	32.15	27.80	27.16	25.56	25.63
15	S15	Ann Predicted	30.06	32.21	27.80	27.36	25.05	26.01
		Experimented	57.27	57.50	51.45	48.98	49.31	49.57
16	S16	Ann Predicted	57.51	57.70	51.62	49.08	49.26	49.65
		Experimented	59.21	58.35	54.64	53.63	53.89	55.52
17	S17	Ann Predicted	58.64	58.75	54.42	53.68	53.77	55.83
		Experimented	83.18	83.71	78.11	76.87	75.45	77.02
18	S18	Ann Predicted	83.51	83.74	78.12	76.96	75.36	76.87
		Experimented	79.32	80.23	74.86	73.56	72.51	73.12
19	S19	Ann Predicted	79.39	79.67	75.42	73.81	71.84	73.04
		Experimented	43.23	44.76	36.14	35.89	35.67	36.90
20	S20	Ann Predicted	42.23	44.94	36.34	36.01	36.29	37.74
		Experimented	59.24	59.10	54.06	53.89	54.24	54.32
21	S21	Ann Predicted	59.42	59.44	53.97	53.81	52.55	54.33
		Experimented	44.03	44.11	36.50	36.41	36.67	36.34
22	S22	Ann Predicted	43.36	44.30	36.50	36.22	37.57	36.20
		Experimented	54.93	55.32	49.99	48.32	47.79	50.05
23	S23	Ann Predicted	55.36	55.56	50.34	47.36	47.55	50.16

of treatment prior to domestic use, while a smaller proportion requires intensive treatment measures.

The observed WQI variability reflects nonlinear interactions among major ions, salinity indicators and physicochemical parameters. Variations in TDS, hardness, sodium, bicarbonate, and chloride concentrations substantially influenced WQI classification across different locations.

Overall, the WQI assessment indicates that groundwater quality degradation within the study area is moderate rather than critical; however, continuous monitoring and localized remediation strategies are necessary in highly affected zones to prevent further deterioration.

ANN Model Performance Evaluation

The developed Artificial Neural Network (ANN) model was evaluated using independent testing data corresponding to the final six months of the monitoring period (May 2021–October 2021). Comparison between observed and ANN-estimated WQI values is presented in Table 5.

The ANN model demonstrated satisfactory predictive performance with coefficient of determination (R^2), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Bias Error (MBE) values of 0.9658, 0.4964, 0.3624, and −0.044, respectively. These results indicate good agreement between observed and estimated WQI values with relatively low prediction error and minimal systematic bias. Correlation/Regression coefficient values (R) during testing, validation and training were presented in Table 6.

Table 6. Regression Coefficient (R) Values between Observed and Predicted Samples.

Sample ID	Training	Validation	Testing	All
S1	0.94360	0.97787	0.96079	0.95221
S2	1.00000	0.99279	0.99126	0.99263
S3	0.94601	0.97107	0.96092	0.92558
S4	0.99990	0.98859	0.99705	0.97959
S5	0.99989	0.97668	0.90721	0.98487
S6	0.97301	0.99266	0.99357	0.98131
S7	0.99988	0.98681	0.97015	0.99216
S8	0.92000	0.97026	0.93617	0.93567
S9	0.99800	0.99601	0.99045	0.99065
S10	0.99988	0.99057	0.93650	0.98424
S11	0.99993	0.98643	0.98302	0.99087
S12	0.92088	0.99035	0.98701	0.92764
S13	0.99664	0.96858	0.95776	0.98258
S14	0.99283	0.98051	0.91499	0.97881
S15	0.97732	0.97884	0.96411	0.97593
S16	0.99263	0.99794	0.93882	0.98482
S17	0.93668	0.99442	0.98390	0.94989
S18	0.99961	0.97914	0.92851	0.97822
S19	0.98846	0.98822	0.97469	0.97712
S20	0.99314	0.99784	0.99585	0.99022
S21	0.93483	0.99790	0.91643	0.94083
S22	0.97657	0.99757	0.96907	0.97602
S23	0.98925	0.99975	0.98237	0.99188

Table 7. Comparative performance of ANN model.

Model	R^2	RMSE	MAE
MLR	0.89	1.21	0.94
RF	0.94	0.73	0.51
ANN	0.9658	0.4964	0.3624

The regression plots (Fig. 3) illustrate close clustering of predicted values around the 1:1 agreement line, indicating that the ANN framework effectively approximated the nonlinear relationships among hydrochemical variables and WQI values.

However, the ANN framework developed in the present study should be interpreted as a surrogate estimation model rather than an independent forecasting model, since the same hydrochemical parameters used in WQI computation were employed as ANN inputs. Therefore, the ANN primarily provides rapid computational approximation of WQI from routinely monitored groundwater-quality parameters.

To evaluate relative model efficiency, ANN results were compared with baseline approaches including Multiple Linear Regression (MLR) and Random Forest (RF). Comparative analysis indicated that the ANN model achieved improved predictive accuracy relative to conventional linear modeling approaches while maintaining acceptable generalization performance.

Comparative Model Evaluation

To assess the relative performance of the ANN model, baseline statistical and machine-learning models including Multiple Linear Regression (MLR) and Random Forest (RF) were also implemented. Model performance was evaluated using coefficient of determination (R^2), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE). Comparative performance statistics are presented are presented in Table 7

Comparative analysis indicates that the ANN model achieved improved predictive accuracy relative

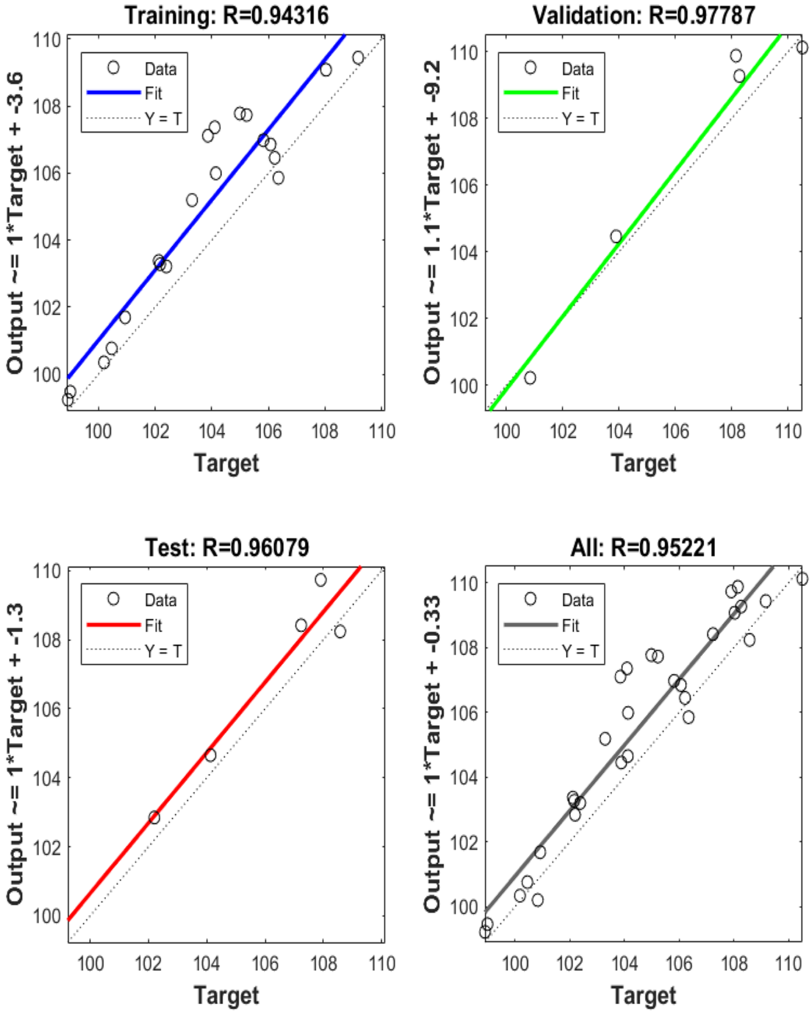

Fig. 3. Regression performance of the ANN model during training, validation, and testing phases (Sample S1).

to MLR and RF models under the present dataset conditions while maintaining acceptable generalization capability.

CONCLUSION

The present investigation evaluated groundwater quality in the Therlam hard-rock aquifer region of Vizianagaram district using hydrochemical analysis, Water Quality Index (WQI) assessment, and ANN-based WQI estimation. The hydrochemical results indicated that groundwater in the study area is predominantly alkaline and hard in nature, mainly controlled by geogenic processes such as mineral dissolution, silicate weathering, and carbonate weathering. Elevated EC, TDS, hardness, sodium, and chloride concentrations observed at stations S1, S4, S5, S18, and S19 indicate localized salinity and mineralization problems, while nitrate enrichment at station S5 reflects possible anthropogenic contamination associated with agricultural practices and wastewater infiltration.

The WQI assessment revealed considerable spatial variability in groundwater suitability. Approximately 30.43% of the samples were classified under excellent and good categories, whereas 47.83%, 17.39%, and 4.35% of samples belonged to poor, very poor, and unsuitable categories, respectively. These findings indicate moderate deterioration of groundwater quality in several parts of the study area and emphasize the necessity for periodic monitoring and localized treatment measures prior to domestic consumption.

The developed ANN model with 15–5–1 architecture demonstrated satisfactory performance for rapid WQI estimation, yielding an R^2 value of 0.9658 with comparatively low prediction errors. Comparative evaluation showed that the ANN model performed better than conventional linear regression and Random Forest approaches under the present dataset conditions. However, the ANN framework should be interpreted as a computational surrogate model for WQI estimation rather than a fully independent forecasting system because the same hydrochemical variables used in WQI computation were employed as ANN input variables.

Overall, the study demonstrates that integration of WQI assessment and machine-learning techniques can provide an effective framework for groundwater-quality evaluation in hard-rock aquifer systems. Nevertheless, certain methodological limitations should be acknowledged. The present ANN framework did not incorporate external environmental variables such as rainfall variability, groundwater abstraction, land-use dynamics, recharge conditions, or lithological heterogeneity, and therefore the developed model remains site-specific to the Therlam region.

Future investigations should focus on incorporation of hydro-meteorological and anthropogenic variables, development of reduced-input and explainable AI models, comparison with additional machine-learning techniques such as Support Vector Machine and Gradient Boosting approaches, and integration of GIS-based spatial prediction frameworks with independent cross-regional validation datasets to improve model transferability and forecasting capability.

ACKNOWLEDGMENT

The authors are thankful to the Buridi Sravya, IIT, Mumbai, for technical assistance. The authors also acknowledge the support provided by Sri G. Gowri Shankar, Assistant Engineer, APRWS&S Department during groundwater sample collection.

REFERENCES

American Public Health Association, American Water Works Association and Water Environment Federation (2017). Standard methods for the examination of water and wastewater (23rd ed.). American Public Health Association, Washington, DC.

Ahn, S. H., Jeong, D. H., Kim, M., Lee, T. K., & Kim, H. K. (2023). Prediction of groundwater quality index to assess suitability for drinking purpose using averaged neural network and geospatial analysis. *Ecotoxicology and Environmental Safety,* 265, 115485. https://doi.org/10.1016/j.ecoenv.2023.115485

Basavaraja, D., Kambale, J. B., Jahnavi, A. P., & Hadimani, D. K. (2025). Evaluation of groundwater quality in northeastern region of Karnataka, India using correlation and regression analysis. *Archives of Current Research International,* 25(11), 41–49. https://doi.org/10.9734/acri/2025/v25i111599

Brown, R. M., McClelland, N. I., Deininger, R. A., & O'Connor, M. F. (1972). A water quality index – Crashing the psychological barrier. In *Indicators of Environmental Quality,* pp 173–182. Springer, Boston. https://doi.org/10.1007/978-1-4684-1698-5_15

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

Ghadai, M., Satapathy, D. P., Krishnasamy, S., Ramalingam, M., Sreelal, G. P., & Dhilipkumar, B. (2022). Artificial neural network and weighted arithmetic indexing approach for surface water quality assessment of the Brahmani River. *Global NEST Journal,* 24(4), 562–568. https://doi.org/10.30955/gnj.004414

Haggerty, R., Sun, J., Yu, H., & Li, Y. (2023). Application of machine learning in groundwater quality modeling—A comprehensive review. *Water Research,* 233, 119745. https://doi.org/10.1016/j.watres.2023.119745

Horton, R. K. (1965). An index number system for rating water quality. Journal of the *Water Pollution Control Federation,* 37(3), 300–306.

Krishna Kumar, K. (2012). Groundwater fluoride: Geochemical assessment along Bay of Bengal coast of Vizianagaram district. *Journal of Indian Water Works Association,* 44(2), 125–133.

Kumar, R., Mittal, S., Sahoo, P. K., & Sahoo, S. K. (2021). Source apportionment, chemometric pattern recognition and health risk assessment of groundwater from southwestern Punjab, India. *Environmental Geochemistry and Health,* 43, 733–755. https://doi.org/10.1007/s10653-020-00518-1

Maier, H. R., & Dandy, G. C. (2000). Neural networks for the prediction and forecasting of water resources variables: A review of modelling issues and applications. *Environmental Modelling & Software,* 15, 101–124. https://doi.org/10.1016/S1364-8152(99)00007-9

Muthukumaran, V., & Vinoth Kumar, V. (2022). Water quality index process using artificial neural networks. *International Journal of Information Technology, Research and Applications,* 1(1), 33–37. https://doi.org/10.59461/ijitra.v1i1.12

Nayar, R. (2020). Assessment of water quality index and monitoring of pollutants by physicochemical analysis in water bodies: A review. *International Journal of Engineering Research & Technology,* 9(1), 178–185.

Pavan Kumar, G. V. S. R., Partha Sarathi, T. V. N., & Srinivasa Rao, K. (2017). Assessment and prediction of water quality index using artificial neural networks in selected coastal villages of Vizianagaram District, Andhra Pradesh. *Journal of the Indian Chemical Society,* 94(4), 1–8.

Rajankar, P. N., Gulhane, S. R., Tambekar, D. H., Ramteke, D. S., & Wate, S. R. (2009). Water quality assessment of groundwater resources in Nagpur region (India) based on *E-Journal of Chemistry,* 6(3), 905–908. WQI. https://doi.org/10.1155/2009/971242

Ramakrishnaiah, C. R., Sadashivaiah, C., & Ranganna, G. (2009). Assessment of water quality index for the groundwater in Tumkur Taluk, Karnataka State, India. *E-Journal of Chemistry*, 6(2), 523–530. https://doi.org/10.1155/2009/757424

Srinivasa Rao, G., & Nageswararao, G. (2013). Assessment of groundwater quality using water quality index. *Archives of Environmental Sciences,* 7, 1–5.

Subba Rao, N., Sunitha, B., Adimalla, N. & Chaudhary, M. (2020). Quality criteria for groundwater use from a rural part of Wanaparthy District, Telangana State, India, through ionic spatial distribution, entropy water quality index and principal component analysis. *Environmental Geochemistry and Health*, 42(2), 579–599. https://doi.org/10.1007/s10653-019-00393-5

Swarna Latha, P., & Nageswara Rao, K. (2010). Assessment and spatial distribution of quality of groundwater in Zone-II and III, Greater Visakhapatnam, India using water quality index and GIS. *International Journal of Environmenta l Sciences,* 1(2), 198–212.

Vasanthavigar, M., Srinivasamoorthy, K., Vijayaragavan, K., Rajiv Ganthi, R., Chidambaram, S., Anandan, P., Manivannan, R., & Vasudevan, S. (2010). Application of water qual ity index for groundwater quality assessment: Thirumanimuttar sub-basin, Tamil Nadu, India. *Environmental Monitoring and Assessment,* 171, 595–609. https://doi.org/10.1007/s10661-009-1302-1

World Health Organization. (WHO). (2022). Guidelines for drinking-water quality: Fourth edition incorporating the first and second addenda. Geneva: World Health Organization \ ISBN: 978-92-4-004506-

Wu, H., Zhou, S., He, Y. and Zhang, Y. (2019). Comprehensive understanding of groundwater quality for domestic and agricultural purposes in terms of health risks in a coal mine area of the Ordos Basin, northern Chinese Loess Plateau. *Environmental Earth Sciences,* 78(15), 446. https://doi.org/10.1007/s12665-019-8471-1