



# Prediction of water quality index using artificial neural network and multiple linear regression modelling approach in Shivganga River basin, India

A. K. Kadam<sup>1</sup> · V. M. Wagh<sup>2</sup> · A. A. Muley<sup>3</sup> · B. N. Umrikar<sup>4</sup> · R. N. Sankhua<sup>5</sup>

Received: 4 December 2018 / Accepted: 14 March 2019 / Published online: 21 March 2019  
© Springer Nature Switzerland AG 2019

## Abstract

The present study has applied artificial neural network (ANN) and multiple linear regression (MLR) techniques to predict the fitness of groundwater quality for drinking from Shivganga River basin, located on the eastern slopes of the Western Ghat region of India. In view of this, thirty-four (34) representative groundwater samples have been collected and analyzed for major cations and anions during pre- and post-monsoon seasons of 2015. The physicochemical parameters such as pH, EC, TDS, TH, Ca, Mg, Na, K, Cl, HCO<sub>3</sub>, SO<sub>4</sub>, NO<sub>3</sub> and PO<sub>4</sub> were considered for computing water quality index (WQI). Analytical results confirmed that all the parameters are within acceptable range; however, EC, TDS, TH, Ca and Mg are exceeding the desirable limit of the WHO drinking standards. The groundwater suitability for drinking was ascertained by WQI method. The WQI value ranges from 25.75 to 129.07 and from 37.54 to 91.38 in pre- and post-monsoon seasons, respectively. Only one sample (DW5) shows 129.07 WQI value indicating poor quality for drinking due to input of domestic and agricultural waste. In the view of generating consistent and precise model for prediction of WQI-based groundwater quality, a Levenberg–Marquardt three-layer back propagation algorithm was used in ANN architecture. Further, MLR model is used to check the efficiency of ANN prediction. The results corroborated that predictions of ANN model are satisfactory and confirms consistently acceptable performance for both the seasons. The proposed ANN model may be useful in similar studies of groundwater quality prediction for drinking purpose.

**Keywords** ANN · MLR · WQI · WHO · Shivganga River · India

## Introduction

Groundwater is one of the important sources of drinking water supply for the people all over the world; however, its quality is extremely sensitive and crucial problem due to anthropogenic pressures in many countries. The water quality is directly coupled with human health. Due to consumption of contaminated water, the health issues are increasing

globally and mostly lead to hike in rate of morbidity and mortality especially amongst the children and poses serious threat to public health vis-à-vis mass epidemic in developing countries (Panaskar et al. 2016; Mukate et al. 2017; Wagh et al. 2018a, b). Moreover, approximately 250 million people get infected yearly, among them 10–20 million deaths mostly occur in developing nations (Dzwaïro et al. 2006). In recent years, several researchers proved that groundwater quality

✉ V. M. Wagh  
wagh.vasant@gmail.com

A. K. Kadam  
kadamajaykumar1@gmail.com

A. A. Muley  
aniket.muley@gmail.com

B. N. Umrikar  
bnumrikar@gmail.com

R. N. Sankhua  
sankhua12@yahoo.com

<sup>1</sup> Department of Environmental Sciences, Savitribai Phule Pune University, Pune 411007, India

<sup>2</sup> School of Earth Sciences, Swami Ramanand Teerth Marathwada University, Nanded, Maharashtra, India

<sup>3</sup> School of Mathematical Sciences, Swami Ramanand Teerth Marathwada University, Nanded, Maharashtra, India

<sup>4</sup> Department of Geology, Savitribai Phule Pune University, Pune 411007, India

<sup>5</sup> Basin Planning, Central Water Commission, New Delhi 110066, India

has conspicuously deteriorated in recent years in most of the countries (Jeong 2001; Moon et al. 2004; Adimalla 2018; Gaikwad et al. 2019). Nowadays, due to inadequate fresh water resources, people are extensively using groundwater for drinking, irrigation and industrial purposes. Generally, groundwater is considered to be safe and reliable source of drinking water due to its natural, hidden existence and less vulnerable for contamination as compared with surface water. The quality of groundwater is an account of biotic, biochemical, and physical physiognomies in reference with quality standards for drinking (Khalil et al. 2011; Gazzaz et al. 2012; Wagh et al. 2016a). Therefore, groundwater quality evaluation is very important based on properties such as physical, chemical and biological, with reference to naturally occurring quality, human health impacts, and proposed uses wherein it depends on the amount of rain, rejuvenation of groundwater and water residence time on and within the surface (Logeshkumaran et al. 2014). Thus, monitoring of water quality is mandatory for the better management of accessible resources of water and to build up various remediation strategies in respective regions.

A water quality index (WQI) is a mathematical tool and composite indicator that provides information to be communicated to end users based on selected water content variables, converted into a single unit less value (Brown and Matlock 2011; Horton 1965; Brown et al. 1970). The WQIs have an advantage of determination of water quality status without interpreting the parameters individually. However, more than 20 water quality indices were developed and revived worldwide till 1970 (Bhargava 1983). In view of the easiness of their use and scientific base, several researchers used various WQIs for assessing water quality. In result, a huge water quality data are produced by the analysis and these data need compilation to conclude water quality status in a particular region. The groundwater quality indices have been developed to incorporate a set of parameters to generate a single index. Moreover, it is a dimensionless index that assigns an appropriate categorical value to cumulative set of measured chemical parameters of groundwater (Pesce and Wunderlin 2000). Subsequently, the WQI can be defined as a single numeric score that express groundwater quality status at a meticulous location over an explicit period (Kaurish and Younos 2007). Banerjee et al. (2018) carried out water quality assessment influenced by highway-broadening-induced activities in the Eastern district of Sikkim. Yaseen et al. (2018) studied various hybrid intelligence models of adaptive neuro-fuzzy inference system, integrated with Fuzzy C-Means data clustering; Grid Partition and Subtractive Clustering models to ascertain the quality of river water. Aboodi et al. (2018) assessed the water quality of Shatt Al-Arab River and its appropriateness for different needs near the Hartha and Najibia power plants through WQI method, organic pollution index, etc., during the summer

and monsoon seasons. Miladenovic Ranisavljevic and Zera-jic (2018) have used various models to compare WQI in the assessment of Danube River surface water quality in the Serbia. Moreover, Shooshtarian et al. (2018) used groundwater quality index based on fuzzy multi-criteria group decision-making models to monitor the impact of land-use changes on groundwater quality in the Iran.

The ANN method is a form of artificial intelligence, which by means of their architecture attempt to simulate and imitate the biological structure of the human brain and nervous system. A neural network consists of simple synchronous processing elements called as neurons that are inspired by biological nerve system (Malinova and Guo 2004). The general ANN network is composed of input, hidden or middle and output layers. In Multilayer Perceptron, the error function is alternative to the weighting coefficients. The ANN model requires high accuracy in technical components for design, development and expansion. Therefore, the input raw data are standardized and optimized for the prevention of the excessive declination in assigned weights. The normalized data are used to increase the processing speed and accuracy of the ANN's performance. ANNs are usually a structure finalized by the designer and input data weights are automatically trained using an optimization algorithm like back propagation and Levenberg–Marquardt optimization algorithms (Huang and Huang 1990). Barzegar and Moghaddam (2016) investigated and compared the accuracy of three different neural network computing techniques viz., multilayer perceptron, radial bias and generalized regression in prediction of groundwater salinity of Tabriz plain (Barzegar and Asghari Moghaddam 2016). Further, three models were combined for improvement in the accuracy of target prediction using committee neural network. Darbandi and Pourhosseini (2018) studied the input combination of the models using Gamma test, MLP–ANN and hybrid multilayer perceptron (MLP–FFA) was used to forecast monthly river flow for a set of time intervals using observed data. Kiraz et al. (2018) used ANN model to predict the efficiency of Pb (II) adsorption and best results were obtained with ten neurons and proved that such a model could save time and cost. In addition, artificial intelligence-based models were used for the prediction of water quality parameters of the Karoon River, Iran by Emamgholizadeh et al. (2014). The efficiency of these models was evaluated based on coefficient of determination ( $R^2$ ), root mean square error (RMSE) and mean absolute error (MAE).

Yilma et al. (2018) developed ANN model and demonstrated its appropriateness for the prediction of CCME WQI in Akaki region of the Ethiopia. Salari et al. (2018) used ANN model for the quality assessment and characterization of physical and chemical parameters of potable water. CCME WQI model was used to assess the groundwater suitability for drinking and irrigation purposes in the Kadava

river basin of Maharashtra (Wagh et al. 2017a). Moreover, ANN model is used to predicate the groundwater suitability for irrigation by considering irrigation indices in the Nanded tehsil of Maharashtra state (Wagh et al. 2016b). The Levenberg–Marquardt three-layer back propagation, traditional back propagation algorithm, resilient back propagation with and without weight algorithms, etc., were used to develop ANN model for prediction of nitrate content in groundwater of Kadava River basin of Nashik district of Maharashtra state (Wagh et al. 2017b, 2018c). Apart from this, in recent times, many researchers have been used ANN models in hydrology and soil-related studies in their respective regions (Hassan et al. 2018; Sihag 2018; Javdanian 2017; Sakizadeh 2016; Sharma et al. 2015).

The explicit use of ANN to develop projecting model for the prediction of WQI for Shivganga watershed is an application which has yet to be investigated (Kadam et al. 2017, 2018). Therefore, the present study is initiated by setting the objectives (i) to evaluate the groundwater quality for drinking by calculating WQI, (ii) to develop ANN and MLR models for prediction of WQI, and (iii) to compare ANN and MLR models to find the precise values of WQI.

### Study area

The present study covers an area of 176 km<sup>2</sup> located on the Eastern slopes of the Western Ghat region of Maharashtra state, India (Fig. 1). It is included in Survey of India (SoI) toposheet numbers 47F/15 and 47F/16 on 1:50,000 scale and lies between 73°44'1.131"E and 73°56'17.941"E longitude and 18°13'36.059"N–18°24'7.466"N latitudes. The proposed area is drained by River Shivganga which is a fifth order stream. The river flows towards the East for about 8 km and then abruptly changes its course to the South, having river total course of 27 km ultimately empties its water into the river Gunjawani. Shivganga watershed features a tropical monsoon climate. The study area receives maximum rain from the southwest monsoonal wind (June and September) and average annual precipitation is about 750 mm. The annual temperature ranges from 39 to 10 °C during summer and winter seasons. The study area includes 58 villages having around 0.07 million population. The area under investigation is bounded by high hill ranges made up of Deccan trap basaltic flows. The hill ranges show the presence of a number of peaks of more than 1000-m altitude.

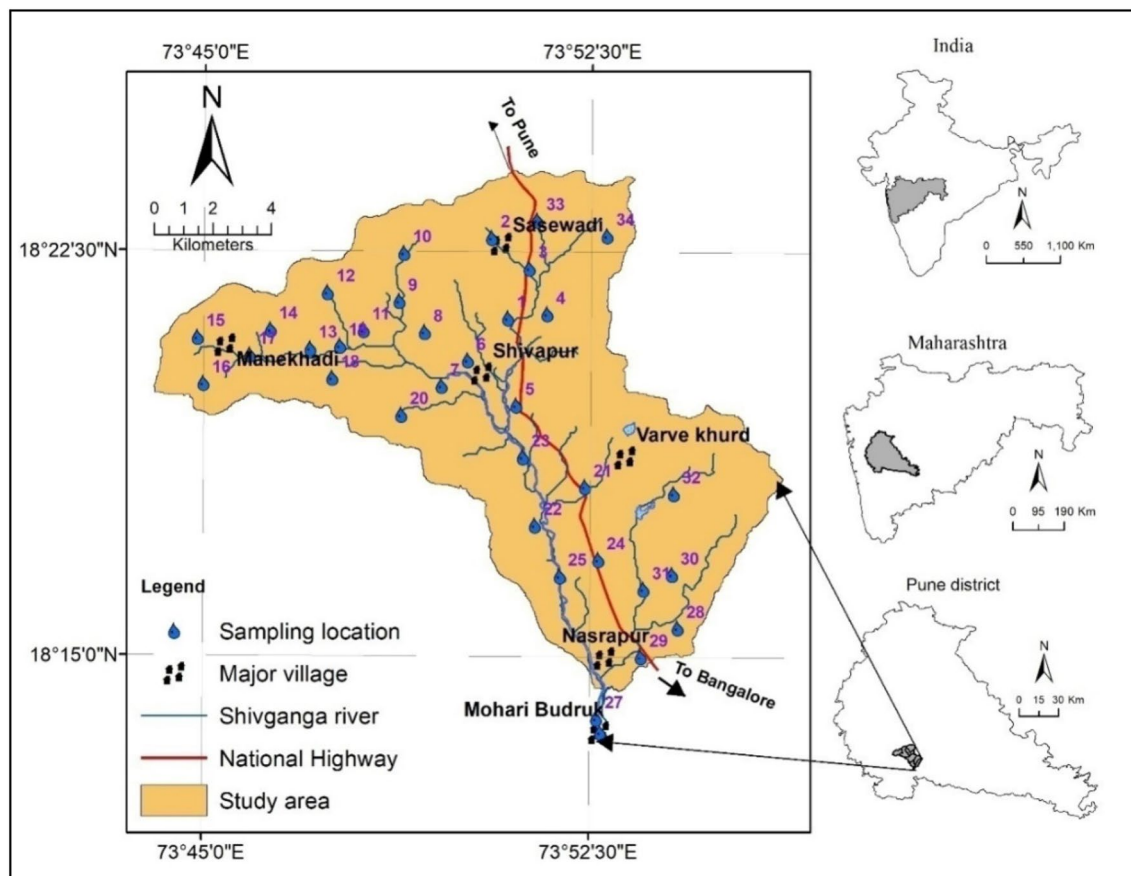


Fig. 1 Shivganga watershed representing groundwater sample locations

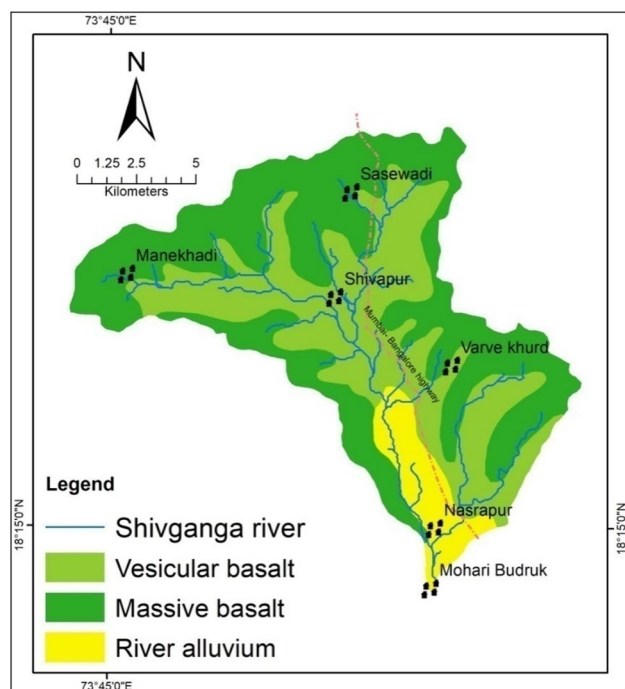
The highest peak (1316 m) is observed at Singhgarh plateau; whereas, the lowest point of 610 m is noted at the southernmost tip of the basin near the confluence of the river. About 30% area is lying above 910 m comprising of a hilly terrain. Moreover, around 16% area is lying between 835 and 910 m which possesses gently rolling topography; however, about 54% area lying below 835 m altitude is largely occupied by flat surface, which denotes a plane land of considerable areal extent with negligible relief (Kadam et al. 2018).

The area under investigation is a part of Deccan Volcanic Province (DVP) comprising of poor to moderately weathered basaltic flows of Wai Subgroup. These basaltic flows exhibit multiple aquifers that are separated by thin impermeable tuffaceous layers known as red boles. The Wai Subgroup consists of five formations viz., Poladpur, Ambenali, Mahabaleshwar, Panhala and Desur are hierarchically from base to top, respectively. In addition, these formations are detached by marker Giant Plagioclase Basalts (GPBs) (Beane et al. 1986). The hydrogeological map (Fig. 2) of the study area is prepared by mapping the segments of streams, road cuttings and Ghat divisions with the help of GSI map (scale 1:250,000). Generally, groundwater in the area occurs within the weathered, amygdaloidal/vesicular, and jointed/fractured compact basaltic aquifers. The groundwater occurrence and availability is influenced by the amount of weathering, jointing, presence of interconnected vesicles and connection of vesicles with fissures and cracks. In the study area, thickness of weathered zone varies broadly up to 10 m bgl, even so, weathered and fractured trap exists in the topographic lows which appear to

be the potential aquifer. Generally, the shallower weathered zones up to the depth of 18–22 m bgl form the phreatic aquifer. Alluvial patches along the river banks, flood plain areas and along second and third order streams are observed.

## Materials and methodology

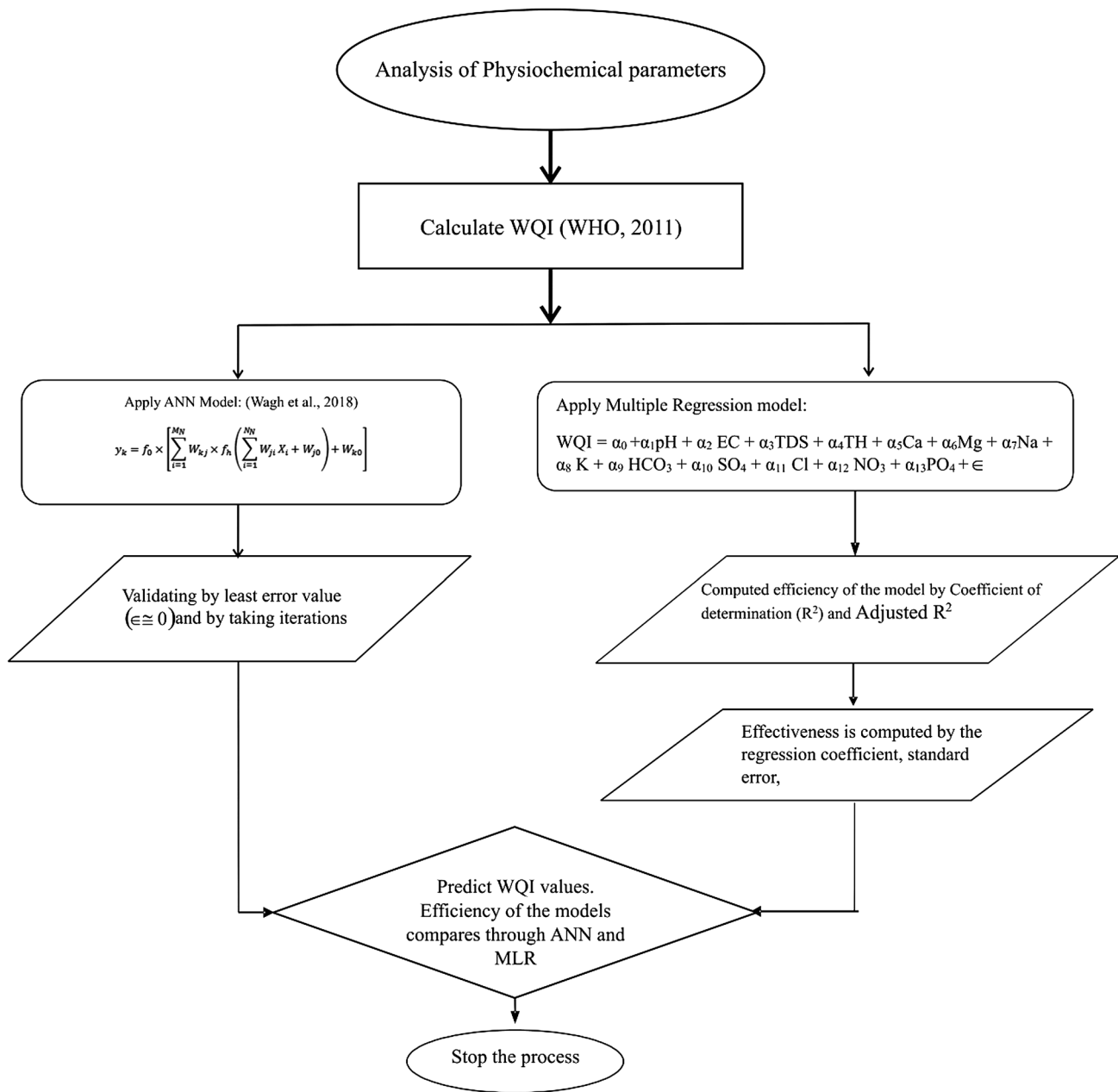
To ascertain the seasonal variations in the groundwater quality, a total 34 dug well samples were collected and analyzed for major ions from Shivganga watershed during pre-monsoon (May 2015) and post-monsoon (November 2015) seasons (Fig. 1). In this context, a random sampling method was employed for the collection of groundwater sample with due consideration to represent various land-use patterns, geomorphology and topography of the study area. The samples were collected in pre-cleaned polyethylene container of 1-L capacity. pH and EC were recorded in situ with handheld digital pH and EC meter. The sample location coordinates were marked using GPS (Garmin) and further exported to GIS software for preparation of the base map of the study area. The collected groundwater samples were brought to the laboratory and kept in refrigerator at 4 °C temperature until analyzed by following standard procedures of American Public Health Association (APHA 2005). The total hardness (TH), calcium ( $\text{Ca}^{2+}$ ), magnesium ( $\text{Mg}^{++}$ ) and chloride (Silver nitrate method) were determined by titrimetric procedures. Sodium ( $\text{Na}^+$ ) and potassium ( $\text{K}^+$ ) contents were analyzed using flame photometer (Systronic-130 model). While sulphate ( $\text{SO}_4^{--}$ ) by turbidometry method, phosphate ( $\text{PO}_4^{--}$ ) by  $\text{SnCl}_2$  method, and nitrate ( $\text{NO}_3^-$ ) by Brucine sulphate method were analyzed using UV–VIS double beam spectrophotometer. The groundwater quality for drinking was assessed by calculating WQI values for all groundwater samples by referring WHO (2011) drinking water standards. Thereafter, WQI prediction was made using artificial neural network (ANN) with training (70%) and testing (30%) data. To authenticate the precision and optimality of ANN model, the least error method ( $\epsilon \cong 0$ ) has been selected. Here, Levenberg–Marquardt three-layer back propagation algorithm is operated for prediction of WQI. Further, multiple regression analysis is performed to check the efficiency of ANN model. R 3.3.3v software is used to analyze the data with the various library functions viz., nnet, quantmod, devtools, NeuralNet tools and metrics. The methodology adopted for development of WQI, MLR and ANN model is shown in the flowchart (Fig. 3).



**Fig. 2** Hydrogeological map of Shivganga River

## Calculation of WQI

Water quality index is helpful method for reflecting the composite weakness of water quality. In addition, it assists in characterizing the water quality to mark potable issues and



**Fig. 3** Flowchart of architecture of ANN and MLR models for WQI prediction

enhances the convenience of protecting activities. The WQI was computed by assigning weight ( $w_i$ ) to each physicochemical parameter based on their relative significance in drinking water. In the present study, the physicochemical parameters viz., pH, EC, TDS, TH, Ca, Mg, Na, K, Cl,  $HCO_3$ ,  $SO_4$ ,  $NO_3$  and  $PO_4$  were considered for calculation of WQI. Initially, each physicochemical parameter was assigned with the weight ( $w_i$ ) in a scale of 1–5 based on its importance in human health and drinking suitability. Thus, maximum weight of 5 was assigned to TDS,  $SO_4$ ,  $NO_3$  and Cl owed to their considerable

significance in drinking water quality. However, PH and EC were allotted by weight 4; TH, Ca, Mg with 3; Na and K as 2 and  $HCO_3$  and  $PO_4$  were given minimum weight of 1 due to the least importance in drinking water fitness (Yidana et al. 2010; Varol and Davraz 2014; Wagh et al. 2017c; Vasant et al. 2016; Şener et al. 2017). The WQI range and type of water have been classified and represented in Table 1. The statistical database of physicochemical parameters has been given in Table 2. The summary of WHO (2011) drinking standards, assigned weight ( $w_i$ ) and relative weight ( $W_i$ ) of



each physicochemical parameter are illustrated in Table 3. The relative weight ( $W_i$ ) is calculated using the following equation:

$$W_i = \frac{w_i}{\sum_{i=1}^n w_i}, \tag{1}$$

where  $W_i$  is the relative weight,  $w_i$  is the weight/parameter and  $n$  is the number of parameters.

A quality rating scale ( $q_i$ ) for each parameter is calculated based on the following equation:

$$q_i = (C_i/S_i) \times 100, \tag{2}$$

where  $q_i$  is the quality rating,  $C_i$  is the chemical concentration/water sample (mg/L),  $S_i$  is the WHO drinking water quality standard (mg/L).

$SI_i$  is the sub-index of  $i$ th parameter:

$$SI_i = W_i \times q_i. \tag{3}$$

The WQI is calculated by

$$WQI = \sum SI_i. \tag{4}$$

**Table 3** The assigned weights and relative weight of physicochemical parameters

Parameters	% of samples falling above desirable limit		Weight ( $w_i$ )	Relative weight ( $W_i$ )
	PRM 2015	POM 2015		
pH	0	0	4	0.093
EC	80	67	4	0.093
TDS	23	15	5	0.116
TH	97	97	3	0.070
Ca	17	06	3	0.070
Mg	03	03	3	0.070
Na	0	0	2	0.047
K	0	0	2	0.047
Cl	0	0	5	0.116
HCO <sub>3</sub>	0	0	1	0.023
SO <sub>4</sub>	0	0	5	0.116
NO <sub>3</sub>	0	0	5	0.116
PO <sub>4</sub>	0	0	1	0.023
$\Sigma$			43	1.000

**Table 1** The range and type of water for WQI (Sahu and Sikdar 2008)

WQI range	Water types
< 50	Excellent water
50–100	Good water
100.1–200	Poor water
200.1–300	Very poor water
> 300.1	Water unsuitable for drinking purpose

### Results and discussions

The analytical results show that the pH ranges from 7.44 to 8.38 in pre-monsoon (PRM) and 6.85–7.51 in post-monsoon (POM) samples and are within the permissible limit of WHO standards. The average EC values vary from 589.71 to 635.29  $\mu$ S/cm for POM and PRM seasons, respectively and exceed the WHO desirable limit of 500  $\mu$ S/cm. TDS values range from 196.10 to 832.60 mg/L in PRM and 163.5–590.60 mg/L in POM. It is observed that 23% PRM and 15% POM samples show high TDS concentration than desirable limit. The groundwater having high TDS content is probably from leaching and percolation of salts from the

**Table 2** Descriptive statistical summary of physicochemical parameter

	pH	EC	TDS	TH	Ca	Mg	Na	K	HCO <sub>3</sub>	SO <sub>4</sub>	Cl	NO <sub>3</sub>	PO <sub>4</sub>
PRM 2015													
Max	8.38	1070	832.6	604	120.24	74.07	108.5	7.78	320	151.24	248.5	18.65	0.32
Min	7.44	296	196.1	52	8.02	7.8	9.51	0.2	30	12.32	19.1	6.39	0.02
Average	7.75	635.29	404.8	238.78	52.73	26.8	29.06	1.17	163.22	42.17	77.59	12.52	0.06
Std dev	0.21	185.95	140.57	102.86	24.95	12.72	16.42	1.78	67.31	32.31	42.2	2.7	0.05
WHO limit	6.5–8.5	500	500	100	75	50	200	10	500	250	250	45	0.5
POM 2015													
Max	7.51	980	590.6	384	89.78	54.58	24.38	3.2	360	62.12	85.2	13.95	2.9
Min	6.85	240	163.5	96	12.83	2.92	4.84	0.0	100	3.85	10.2	0.12	0.07
Average	7.25	589.71	382.88	247.89	55.68	23.54	13.68	0.4	239.72	17.9	33.57	3.63	0.19
Std dev	0.14	169.61	102.11	80.53	15.27	10.99	4.59	0.6	61.77	13.87	19.76	3.46	0.49

All parameter values are expressed in mg/L; EC is in  $\mu$ S/cm and pH on scale

soil and certain prevailing anthropogenic inputs. TH is varying from 52 to 604 mg/L in PRM (average 238.78 mg/L) and 96–384 mg/L in POM (average 247.89 mg/l), the average value suggesting increased concentration towards POM owed to dissolution of minerals.

The cationic order is  $Ca^{2+} > Na^{+} > Mg^{2+} > K^{+}$  in PRM; where, rainfall is normal and  $Ca^{2+} > Mg^{2+} > Na^{+} > K^{+}$  in POM season in low rainfall condition. The  $Ca^{2+}$  values show wide variation from 8.02 to 120.24 mg/L and 12.83–79.78 mg/L, with an average value of 52.73 mg/L and 55.68 mg/L in PRM and POM seasons, respectively. These values were varying with monsoon as rainfall decreases in PRM and the number of samples above WHO limit increases. Thus, ~ 17% of the groundwater samples from PRM and only 6% samples of POM period are above the WHO limit (75 mg/L). The significant higher concentration of  $Mg^{2+}$  ranges from 7.80 to 74.07 mg/L in PRM and 2.92–54.58 mg/L in POM season. The mean value of  $Na^{+}$  is 24.94 and 17.91 mg/L in PRM and POM period. The average concentration of  $K^{+}$  value is 1.17 and 0.4 mg/L in PRM and POM seasons.

Anionic order in groundwater samples is  $HCO_3^{-} > SO_4^{2-} > Cl^{-} > NO_3^{-} > PO_4^{2-}$  in PRM and  $HCO_3^{-} > Cl^{-} > SO_4^{2-} > NO_3^{-} > PO_4^{2-}$  in POM seasons. The  $HCO_3^{-}$  concentration varies from 30 to 320 mg/L for PRM and 100–360 mg/L for PRM season. The elevated concentration of  $HCO_3^{-}$  in groundwater samples is owed by the host basaltic rock (Pawar et al. 2008). The average sulphate concentration is 42.17 mg/L and 17.9 mg/L in PRM and POM seasons. The sources of  $SO_4^{2-}$  in groundwater are through dissolution/oxidation of sulphate minerals and anthropogenic inputs. In addition, chloride concentration in PRM samples is 19.1–248.5 mg/L, and POM season is 10.20–85.20 mg/L. The elevated content of  $Cl^{-}$  are symptomatic due to discharge of sewage effluent, decomposition of organic material and runoff from agricultural

areas (Panaskar et al. 2016). The  $NO_3^{-}$  concentration ranges between 6.39 and 18.65 mg/L during PRM (average of 12.52 mg/L), and for POM season concentration ranges between 0.12 and 13.95 mg/L, with an average of 3.63 mg/L. It is observed that the high concentration of nitrate found in post-monsoon season owing to nitrogen complex fertilizers and percolation of surface water in certain wells due to reprehensible sealing of the dug well walls (Wagh et al. 2019). The  $PO_4$  values range from 0.02 to 0.32 mg/L in PRM and 0.07–2.90 mg/L in POM. The higher values in POM are mainly from agriculture return flow.

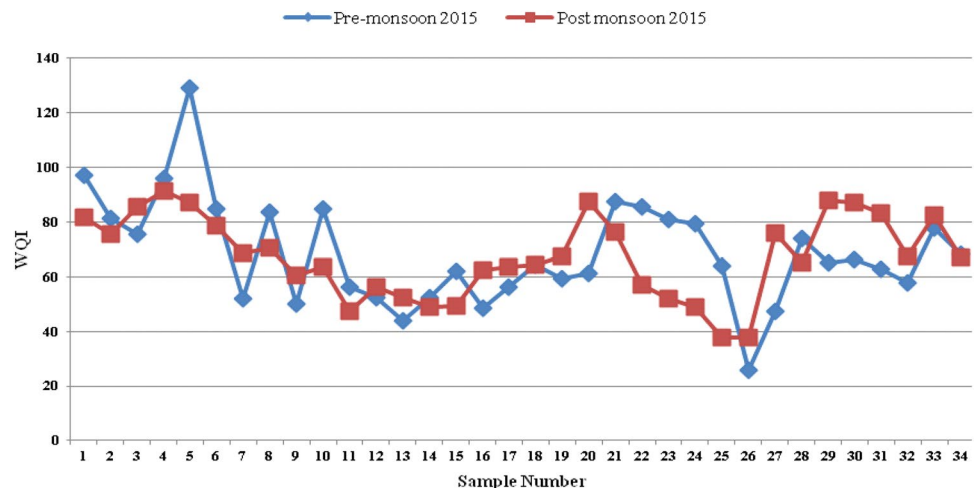
### Assessment of the water quality using WQI

To ascertain the groundwater quality for drinking, WQI was calculated using Eqs. (1)–(4). The WQI values range from 25.75 to 129.07 and 37.54–91.38 in pre- and post-monsoon seasons, respectively. From Fig. 4, it is observed that only DW5 shows 129.07 WQI value indicating poor groundwater quality for drinking purpose mainly due to inputs from domestic and/or agricultural discharge. It is confirmed that groundwater samples from PRM and POM seasons are demonstrating excellent to good quality of groundwater for drinking purpose.

### Artificial neural network model (ANN)

The applicability of ANN was investigated to forecast WQI values in 34 dug wells from the study area. The performance results of the model with LM algorithm are provided in Table 4. The performance of the back propagation LM algorithm was evaluated by monitoring the error between modelled output and measured dataset. The number of neurons was optimized by keeping all other parameters constant

**Fig. 4** Plot of WQI values for pre- and post-monsoon seasons of 2015



**Table 4** Iteration results of the proposed ANN model

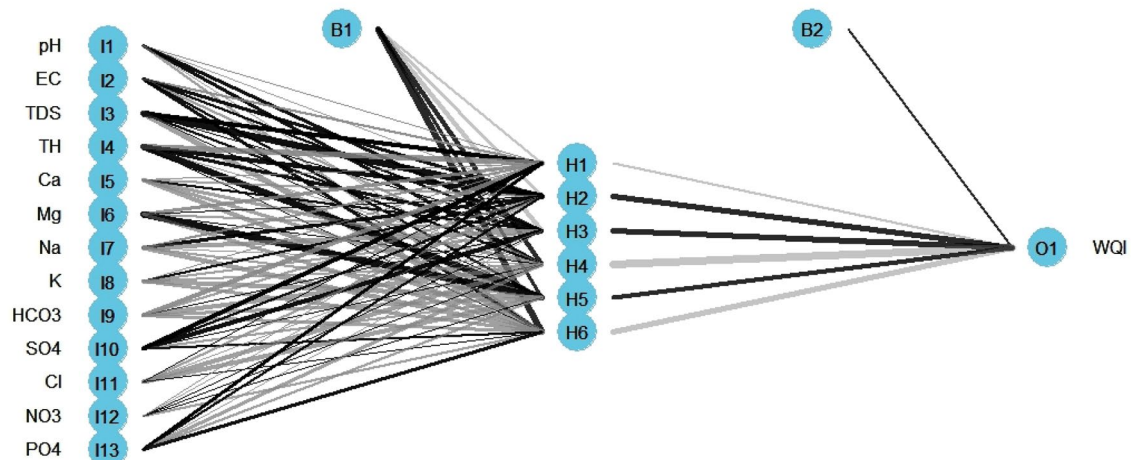
Pre-monsoon 2015		Post-monsoon 2015	
Iteration	Error	Iteration	Error
Initial value	1.052115	Initial value	9.062641
10	0.008977	10	0.034923
20	0.000985	20	0.003912
30	0.000335	30	0.000751
40	0.000176	40	0.000546
Final value	0.000086	50	0.000277
		60	0.000159
		Final value	0.000096

for the output variable, i.e. WQI. The error decreased for the dataset with the appropriate number of hidden neurons. Here, after simulation the optimum 6 numbers of hidden neurons, it is found that the error does not change significantly in pre- and post-monsoon seasons. The optimal ANN structure contained 13 input variables with six hidden neuron and WQI as the output variable. Figures 4 and 5 represent the optimal neural network in pre- and post-monsoon seasons. The  $B_1$  and  $B_2$  are the bias occurred in the model and the number of iterations is helpful in its removal. Based on our proposed approach, the training was stopped at 40th and 60th iteration as error did not change significantly. Finally, at the 40th iteration the least error, i.e. 0.000086 was obtained. In the pre-monsoon season, the initial error (1.05) was high but after taking the number of iterations it has been lowered. Moreover, in the post-monsoon season, the initial error displayed was high, i.e. 9.06; but later on, after taking the number of iterations at the 60th iteration the least error, i.e. 0.000096 (Table 4). The obtained results reveal that initially the generated weight of the applied model is 91 in pre- and post-monsoon season (Figs. 5, 6). The variations

measured are so diminutive; as a result, it proves as an effective tool in the assessment of WQI.

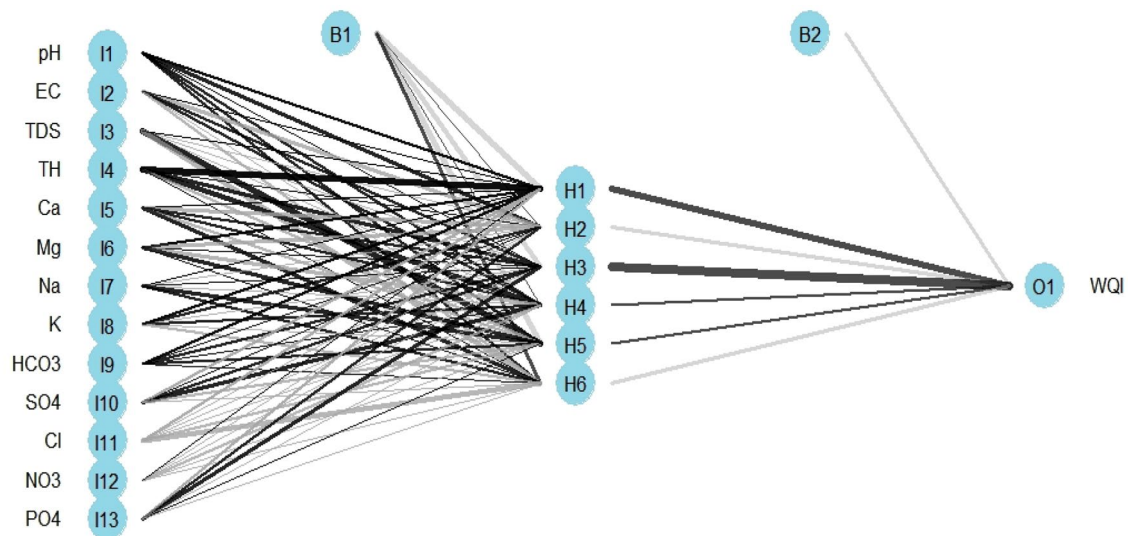
### Multiple linear regression model (MLR)

The MLR model is useful in discovering the association between various independent and dependent variables. The general form of the regression equations is according to flowchart 1. The WQI is a vector representing all location values of WQI, and  $\alpha_{0,\dots}, \alpha_{13}$  are fixed but unknown constants. Also, pH, EC, TDS, TH, Ca, Mg, Na, K,  $\text{HCO}_3$ ,  $\text{SO}_4$ , Cl,  $\text{NO}_3$  and  $\text{PO}_4$  are independent random variables used to foretell WQI by constructing regression equations using R software. The standardized regression coefficient ( $\alpha_j, j=0, 1, 2, \dots, 13$ ; see flowchart 1) computes the change in independent variable and dependent variable, i.e. WQI. The standard error has been used to evaluate the stability of regression coefficients. The  $T$  test is executed to confirm the significance of means of water quality parameters and variance is assessed using  $F$  test (Sahoo and Zha 2013; Beaumont et al. 1984). Vigorous MLR models were assessed by coefficient of determination ( $R^2$ ). The adjusted  $R^2$  is interpreted in the identical way as the ' $R^2$ ' values, except that the adjusted  $R^2$  takes into consideration the number of degrees of freedom. The coefficients of the model and significance of each parameter for pre-monsoon season 2015 are represented in Table 5. The standard error is  $3.545\text{e}-10$  on 20 degrees of freedom (DF), with  $R^2$  is 1 and adjusted  $R^2$  is 1. It is compared through ' $F$ ' value which is  $7.079\text{e}+17$  on 13 and 20 DF, and the ' $p$ ' value is less than  $2.2\text{e}-16$ . It is simply elaborating the efficiency of the present model. However, in post-monsoon season of 2015, the coefficients of the model are represented (Table 6). The standard error is  $3.699\text{e}-10$  on 20 degrees



**Fig. 5** ANN structure for pre-monsoon season of 2015





**Fig. 6** ANN structure for post-monsoon season of 2015

**Table 5** Coefficient of the MLR in pre-monsoon season 2015

Parameter	Coefficient	E (standard error)		t value Pr(>  t )
		Min	Max	
Intercept	- 1.14E-02	3.13E-10	- 3.65E+07	< 2e-16***
pH	1.21E-02	3.93E-10	3.08E+07	< 2e-16***
EC	1.58E-01	5.76E-10	2.74E+08	< 2e-16***
TDS	1.65E-01	3.83E-09	4.31E+07	< 2e-16***
TH	3.74E-01	3.35E-09	1.12E+08	< 2e-16***
Ca	1.01E-01	1.95E-09	5.21E+07	< 2e-16***
Mg	8.98E-02	2.06E-09	4.36E+07	< 2e-16***
Na	2.25E-02	9.62E-10	2.34E+07	< 2e-16***
K	3.45E-02	4.76E-10	7.24E+07	< 2e-16***
HCO <sub>3</sub>	1.17E-01	1.49E-09	7.88E+07	< 2e-16***
SO <sub>4</sub>	5.00E-02	9.34E-10	5.35E+07	< 2e-16***
Cl	- 2.301E-09	1.35E-09	- 1.71E+00	0.103
NO <sub>3</sub>	2.45E-02	3.90E-10	6.28E+07	< 2e-16***
PO <sub>4</sub>	1.34E-06	1.50E-09	8.89E+02	< 2e-16***

Significant codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

**Table 6** Coefficient of the MLR in post-monsoon season 2015

Parameter	Coefficient	E (std error)		t value Pr(>  t )
		Min	Max	
Intercept	- 1.37E-01	8.40E-08	- 1.63E+06	< 2e-16***
pH	1.63E-02	3.80E-10	4.28E+07	< 2e-16***
EC	2.56E-01	7.07E-10	3.61E+08	< 2e-16***
TDS	1.84E-01	1.24E-06	1.48E+05	< 2e-16***
TH	3.74E-01	1.09E-09	3.43E+08	< 2e-16***
Ca	1.33E-01	2.24E-07	5.95E+05	< 2e-16***
Mg	1.34E-01	1.51E-07	8.91E+05	< 2e-16***
Na	8.53E-03	5.70E-08	1.50E+05	< 2e-16***
K	2.78E-02	9.17E-09	3.03E+06	< 2e-16***
HCO <sub>3</sub>	2.02E-01	7.57E-07	2.67E+05	< 2e-16***
SO <sub>4</sub>	4.03E-02	1.70E-07	2.37E+05	< 2e-16***
Cl	3.30E-07	2.18E-07	1.51E+00	0.146
NO <sub>3</sub>	5.31E-02	4.61E-10	1.15E+08	< 2e-16***
PO <sub>4</sub>	2.48E-05	6.91E-10	3.59E+04	< 2e-16***

Significant codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

of freedom, with multiple  $R^2$  is 1 and adjusted  $R^2$  is 1. It is compared through 'F' value which is 1.453e+18 on 13 and 20 DF, and the 'p' value is less than 2.2e-16, elaborates the efficiency of the model.

The predicted WQI values of ANN model gives conformity with the observed values and thus its representativeness is effective in predicting the WQI values. In this study, it is concluded that ANN is appropriate compared to the MLR model. ANN models counterpart convincingly fit quality. MLR modelling technique is based on the simple least square method; whereas, the ANN model imitates the functioning of the human being intelligence. The MLR technique

having significant realistic reward is functioning much simpler and less time consuming (Sahoo and Zha 2013).

### Conclusions

The main focus of this paper is to determine WQI to ascertain the groundwater suitability for drinking. The analytical results authenticated that EC, TDS, TH, Ca and Mg are beyond the desirable limits and rest of parameters are within safe limits of the WHO drinking standards. WQI results inferred that all the groundwater samples fall in good

and excellent category except DW5 showing 129.07 WQI value. Hence, overall water quality in the Shivganga watershed is suitable for drinking. In this study, ANN and MLR models are used to find the accuracy of WQI for future prediction of water quality. The determination of WQI values are validated through ANN and MLR models. The MLR model exhibited that all other variables are found to be significant in the formation of regression model excluding chloride content. The neural network architecture comprising 13 contributing neurons, six hidden neurons and one output variable were used for WQI prediction. The obtained result of ANN model is well-accepted least error ( $\epsilon \cong 0$ ) in pre- and post-monsoon seasons of 2015. The comparison of ANN and MLR models suggests that the precision level is high in ANN model. Hence, the preference is given to ANN model due to its iterative moves to get more accurate results in both the seasons. Therefore, ANN model would become more beneficial in the prediction of water quality in future. Consequently, the MLR model can serve as an alternative and cost-effective tool for groundwater quality prediction in the circumstances, where trained expertise and time constraints and the field data are favourable. The ANN model is most preferred due to restrain time and exertions encumber of repetitive WQI grit, a modelling approach that can be employed fruitfully for adequate results in similar studies. This study recommends that there must be enough protection of water resources to preserve its quality in the present state. In addition, it helps for a long period of time to avoid any possible water contaminant in the future. The proposed ANN model can be further improved by more detailed meteorological and spatially distributed water quality data to exhibit the comprehensiveness of the proposed approach to research communities in the context of developing groundwater management or protection plan.

**Acknowledgements** The authors extend their sincere thanks to Head, Departments of Geology and Environmental Sciences Savitribai Phule, Pune University, Pune for providing necessary facilities for analysis. In addition, thanks to anonym reviewers for meaningful suggestions to enhance the manuscript quality.

## References

- Adimalla N (2018) Groundwater quality for drinking and irrigation purposes and potential health risks assessment: a case study from semi-arid region of South India. *Exposure Health*. <https://doi.org/10.1007/s12403-018-0288-8>
- Al-Aboodi AH, Abbas SA, Ibrahim HT (2018) Effect of Hartha and Najibia power plants on water quality indices of Shatt Al-Arab River, south of Iraq. *Appl Water Sci* 8:64. <https://doi.org/10.1007/s13201-018-0703-0>
- APHA, Federation and WE and American Public Health Association (2005) Standard methods for the examination of water and wastewater 2005. American Public Health Association (APHA), Washington
- Banerjee P, Ghose MK, Pradhan R (2018) AHP-based spatial analysis of water quality impact assessment due to change in vehicular traffic caused by highway broadening in Sikkim Himalaya. *Appl Water Sci* 8:72. <https://doi.org/10.1007/s13201-018-0699-5>
- Barzegar R, Asghari Moghaddam A (2016) Combining the advantages of neural networks using the concept of committee machine in the groundwater salinity prediction. *Model Earth Syst Environ* 2:26. <https://doi.org/10.1007/s40808-015-0072-8>
- Beane JE, Turner CA, Hooper PR et al (1986) Stratigraphy, composition and form of the Deccan Basalts, Western Ghats, India. *Bull Volcanol* 48:61–83. <https://doi.org/10.1007/BF01073513>
- Beaumont C, Makridakis S, Wheelwright SC, McGee VE (1984) Forecasting: methods and applications. *J Oper Res Soc* 35:79. <https://doi.org/10.2307/2581936>
- Bhargava D (1983) Use of water quality index for river classification and zoning of Ganga river. *Environ Pollut Ser B Chem Phys* 6:51–67. [https://doi.org/10.1016/0143-148X\(83\)90029-0](https://doi.org/10.1016/0143-148X(83)90029-0)
- Brown A, Matlock MD (2011) A review of water scarcity indices and methodologies. University of Arkansas. The sustainability consortium, Fayetteville, p 21
- Brown RM, McClelland NI, Deininger RA, Tozer RG (1970) A water quality index-do we dare? water sewage. Works, October, pp 339–343
- Darbandi S, Pourhosseini FA (2018) River flow simulation using a multilayer perceptron-firefly algorithm model. *Appl Water Sci* 8:85. <https://doi.org/10.1007/s13201-018-0713-y>
- Dzwairo B, Hoko Z, Love D, Guzha E (2006) Assessment of the impacts of pit latrines on groundwater quality in rural areas: a case study from Marondera district. *Zimbabwe Phys Chem Earth* 31:779–788. <https://doi.org/10.1016/j.pce.2006.08.031>
- Emamgholizadeh S, Kashi H, Marofpoor I, Zalaghi E (2014) Prediction of water quality parameters of Karoon River (Iran) by artificial intelligence-based models. *Int J Environ Sci Technol* 11(3):645–656
- Gaikwad S, Gaikwad S, Meshram D, Wagh V, Kandekar A, Kadam A (2019) Geochemical mobility of ions in groundwater from the tropical western coast of Maharashtra, India: implication to groundwater quality. *Environ Dev Sustain* 1–34
- Gazzaz NM, Yusoff MK, Aris AZ et al (2012) Artificial neural network modeling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors. *Mar Pollut Bull* 64:2409–2420. <https://doi.org/10.1016/j.marpolbul.2012.08.005>
- Hassan M, Zaffar H, Mehmood I, Khitab A (2018) Development of streamflow prediction models for a weir using ANN and step-wise regression. *Model Earth Syst Environ* 4(3):1021–1028
- Horton RK (1965) An index number system for rating water quality. *J Water Pollut Control Fed* 37(3):300–306
- Huang S-C, Huang Y-F (1990) Learning algorithms for perceptions using back-propagation with selective updates. *IEEE Control Syst Mag*. <https://doi.org/10.1109/37.55125>
- Javdanian H (2017) Assessment of shear stiffness ratio of cohesionless soils using neural modeling. *Model Earth Syst Environ* 3(3):1045–1053
- Jeong CH (2001) Effect of land use and urbanization on hydrochemistry and contamination of groundwater from Taejon area. *Korea J Hydrol* 253:194–210. [https://doi.org/10.1016/S0022-1694\(01\)00481-4](https://doi.org/10.1016/S0022-1694(01)00481-4)
- Kadam AK, Jaweed TH, Umrikar BN et al (2017) Morphometric prioritization of semi-arid watershed for plant growth potential using GIS technique. *Model Earth Syst Environ* 3:1663–1673. <https://doi.org/10.1007/s40808-017-0386-9>
- Kadam A, Karnewar AS, Umrikar B, Sankhua RN (2018) Hydrological response-based watershed prioritization in semiarid, basaltic region of western India using frequency ratio, fuzzy logic and AHP method. *Environ Dev Sustain*

- Kaurish FW, Younos T (2007) Developing a standardized water quality index for evaluating surface water quality. *J Am Water Resour Assoc* 43:533–545. <https://doi.org/10.1111/j.1752-1688.2007.00042.x>
- Khalil B, Ouarda TBMJ, St-Hilaire A (2011) Estimation of water quality characteristics at ungauged sites using artificial neural networks and canonical correlation analysis. *J Hydrol* 405:277–287. <https://doi.org/10.1016/j.jhydrol.2011.05.024>
- Kiraz A, Canpolat O, Erkan EF, Özer Ç (2018) Artificial neural networks modeling for the prediction of Pb (II) adsorption. *Int J Environ Sci Technol* 1–8
- Logeshkumaran A, Magesh NS, Godson PS, Chandrasekar N (2014) Hydro-geochemistry and application of water quality index (WQI) for groundwater quality assessment, Anna Nagar, part of Chennai City, Tamil Nadu, India. <https://doi.org/10.1007/s13201-014-0196-4>
- Malinova T, Guo ZX (2004) Artificial neural network modelling of hydrogen storage properties of Mg-based alloys. *Mater Sci Eng A* 365:219–227. <https://doi.org/10.1016/j.msea.2003.09.031>
- Moon SK, Woo NC, Lee KS (2004) Statistical analysis of hydrographs and water-table fluctuation to estimate groundwater recharge. *J Hydrol* 292(1–4):198–209
- Mladenović-Ranisavljević II, Žerajić SA (2017) Comparison of different models of water quality index in the assessment of surface water quality. *Int J Environ Sci Technol*. <https://doi.org/10.1007/s13762-017-1426-8>
- Mukate S, Panaskar D, Wagh V et al (2017) Impact of anthropogenic inputs on water quality in Chincholi industrial area of Solapur, Maharashtra, India. *Groundw Sustain Dev*. <https://doi.org/10.1016/j.gsd.2017.11.001>
- Panaskar DB, Wagh VM, Muley AA et al (2016) Evaluating groundwater suitability for the domestic, irrigation, and industrial purposes in Nanded Tehsil, Maharashtra, India, using GIS and statistics. *Arab J Geosci*. <https://doi.org/10.1007/s12517-016-2641-1>
- Pawar NJ, Pawar JB, Kumar S, Supekar A (2008) Geochemical eccentricity of ground water allied to weathering of basalts from the Deccan Volcanic Province, India: insinuation on CO<sub>2</sub> consumption. *Aquat Geochem* 14(1):41–71
- Pesce SF, Wunderlin DA (2000) Use of Water Quality Indices To Verify the *Â Rdo*ba City (Argentina) on Suquôâ a Impact of Co. *Science* 80: 34
- Sahu P, Sikdar PK (2008) Hydrochemical framework of the aquifer in and around East Kolkata Wetlands, West Bengal, India. *Environ Geol* 55(4):823–835
- Sahoo S, Jha MK (2013) Groundwater-level prediction using multiple linear regression and artificial neural network techniques: a comparative assessment. *Hydrogeol J* 21(8):1865–1887
- Sakizadeh M (2016) Artificial intelligence for the prediction of water quality index in groundwater systems. *Model Earth Syst Environ* 2(1):8
- Salari M, Salami Shahid E, Afzali SH, Ehteshami M, Conti GO, Derakhshan Z, Sheibani SN (2018) Quality assessment and artificial neural networks modeling for characterization of chemical and physical parameters of potable water. *Food Chem Toxicol* 118:212–219. <https://doi.org/10.1016/j.fct.2018.04.036>
- Şener Ş, Şener E, Davraz A (2017) Evaluation of water quality using water quality index (WQI) method and GIS in Aksu River (SW-Turkey). *Sci Total Environ* 584–585:131–144. <https://doi.org/10.1016/j.scitotenv.2017.01.102>
- Sharma N, Zakauallah M, Tiwari H, Kumar D (2015) Runoff and sediment yield modeling using ANN and support vector machines: a case study from Nepal watershed. *Model Earth Syst Environ* 1(3):23
- Shooshtarian MR, Dehghani M, Margherita F, Gea OC, Mortezaazadeh S (2018) Land use change and conversion effects on ground water quality trends: an integration of land change modeler in GIS and a new ground water quality index developed by fuzzy multi-criteria group decision-making models. *Food Chem Toxicol* 114:204–214
- Sihag P (2018) Prediction of unsaturated hydraulic conductivity using fuzzy logic and artificial neural network. *Model Earth Syst Environ* 4(1):189–198
- Varol S, Davraz A (2014) Evaluation of the groundwater quality with WQI (water quality index) and multivariate analysis: a case study of the Tefenni plain (Burdur/Turkey). *Environ Earth Sci* 73:1725–1744. <https://doi.org/10.1007/s12665-014-3531-z>
- Vasant W, Dipak P, Aniket M, Ranjitsinh P, Shrikant M, Nitin D, Manesh A, Abhay V (2016) GIS and statistical approach to assess the groundwater quality of Nanded Tehsil, (MS) India. In: *Proceedings of first international conference on information and communication technology for intelligent systems*, vol 1, pp 409–417. Springer, Cham
- Wagh VM, Panaskar DB, Varade AM et al (2016a) Major ion chemistry and quality assessment of the groundwater resources of Nanded tehsil, a part of southeast Deccan Volcanic Province, Maharashtra, India. *Environ Earth Sci*. <https://doi.org/10.1007/s12665-016-6212-2>
- Wagh VM, Panaskar DB, Muley AA et al (2016b) Prediction of groundwater suitability for irrigation using artificial neural network model: a case study of Nanded tehsil, Maharashtra, India. *Model Earth Syst Environ* 2:196. <https://doi.org/10.1007/s40808-016-0250-3>
- Wagh VM, Panaskar DB, Muley AA, Mukate SV (2017a) Groundwater suitability evaluation by CCME WQI model for Kadava River Basin, Nashik, Maharashtra, India. *Model Earth Syst Environ* 3:557–565. <https://doi.org/10.1007/s40808-017-0316-x>
- Wagh VM, Panaskar DB, Muley AA (2017b) Estimation of nitrate concentration in groundwater of Kadava river basin-Nashik district, Maharashtra, India by using artificial neural network model. *Model Earth Syst Environ* 3(1):36
- Wagh V, Panaskar D, Mukate S, Lolage Y, Muley A (2017c) Groundwater quality evaluation by physicochemical characterization and water quality index for Nanded Tehsil, Maharashtra, India. In: *Proceedings of the 7th international conference on biology, environment and chemistry, IPCBEE*, vol 98
- Wagh VM, Panaskar DB, Mukate SV, Gaikwad SK, Muley AA, Varade AM (2018a) Health risk assessment of heavy metal contamination in groundwater of Kadava River Basin, Nashik, India. *Model Earth Syst Environ* 4(3):969–980
- Wagh V, Panaskar D, Aamalawar M, Lolage Y, Mukate S, Adimalla N (2018b) Hydrochemical characterisation and groundwater suitability for drinking and irrigation uses in semiarid region of Nashik, Maharashtra, India. *Hydrospat Anal* 2 (1): 43–60. <https://doi.org/10.21523/gcj3.18020104>
- Wagh V, Panaskar D, Muley A et al (2018c) Neural network modelling for nitrate concentration in groundwater of Kadava River basin, Nashik, Maharashtra, India. *Groundw Sustain Dev*. <https://doi.org/10.1016/j.gsd.2017.12.012>
- Wagh VM, Panaskar DB, Mukate SV, Aamalawar ML, Sahu L, U (2019) Nitrate associated health risks from groundwater of Kadava River Basin Nashik, Maharashtra, India. *Hum Ecol Risk Assess* 1–19
- Wan Abdul Ghani WMH, Abas Kutty A, Mahazar MA et al (2018) Performance of biotic indices in comparison to chemical-based water quality index (WQI) in evaluating the water quality of urban river. *Environ Monit Assess*. <https://doi.org/10.1007/s10661-018-6675-6>
- WHO (2011) Guidelines for drinking-water quality, 4th edn. [http://www.who.int/publications/2011/9789241548151\\_eng.pdf](http://www.who.int/publications/2011/9789241548151_eng.pdf)
- Yaseen ZM, Fu M, Wang C et al (2018) Application of the hybrid artificial neural network coupled with rolling mechanism and grey model algorithms for streamflow forecasting over multiple

time horizons. *Water Resour Manag* 32:1883–1899. <https://doi.org/10.1007/s11269-018-1909-5>

Yidana SM, Banoeng-Yakubo B, Akabzaa TM (2010) Analysis of groundwater quality using multivariate and spatial analyses in the Keta basin, Ghana. *J Afr Earth Sci* 58:220–234. <https://doi.org/10.1016/j.jafrearsci.2010.03.003>

Yilma M, Kiflie Z, Windsperger A, Gessese N (2018) Application of artificial neural network in water quality index prediction: a case study in Little Akaki River, Addis Ababa, Ethiopia. *Model*

*Earth Syst Environ* 4:175–187. <https://doi.org/10.1007/s40808-018-0437-x>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.