



Estimation of groundwater quality using an integration of water quality index, artificial intelligence methods and GIS: Case study, Central Mediterranean Region of Turkey

Sevda Taşan¹

Received: 7 July 2021 / Accepted: 31 October 2022 / Published online: 21 November 2022
© The Author(s) 2022

Abstract

Groundwater is one of the most important natural resources in the world and is widely used for irrigation purposes. Groundwater quality is affected by various natural heterogeneities and anthropogenic activities. Consequently, monitoring groundwater quality and assessing its suitability are crucial for sustainable agricultural irrigation. In this study, the suitability of groundwater for irrigation was determined by using sodium adsorption ratio (SAR), residual sodium carbonate (RSC), Kelly index (KI), percentage of sodium (Na%), magnesium ratio (MR), potential salinity (PS) and permeability index (PI). The groundwater samples were collected and analyzed from 37 different sampling stations for this purpose. Along with suitability analysis, artificial neural network (ANN) and adaptive neuro-fuzzy inference system (ANFIS) models were used to predict irrigation water quality parameters. The models were evaluated by comparing the measured values and the predicted values using the statistical criteria [coefficient of determination (R^2), mean absolute error (MAE), root mean square error (RMSE) and Nash–Sutcliffe efficiency (NS)]. In the estimation of all irrigation water quality parameters, the ANN model has performed much higher compared with the ANFIS model. Spatial distribution maps were generated for measured and ANN model-estimated irrigation water quality indices using the IDW interpolation method. Spatial distributions of groundwater quality indices revealed that MR was higher than the allowable limits in most of the study areas and the other quality criteria were within the permissible limits. It has been determined that the interpolation maps obtained as a result of artificial intelligence methods have appropriate sensitivity when compared with the observed maps. Based on the present findings, ANN models could be used as an efficient tool for estimating groundwater quality indices in unsampled sections of the study area and the other regions with similar conditions.

Keywords Artificial neural networks · Adaptive network-based fuzzy inference system · Irrigation suitability · Principal component analysis · Spatial distribution

Introduction

Groundwater is considered as the primary source of water in case of proper or insufficient surface water supply. For this reason, groundwater quality has emerged as one of the most critical environmental problems encountered in the last few decades (Ravichandra and Chandana 2006). Groundwaters are used for domestic purposes and as an irrigation water

resource. Therefore, groundwater quality should be determined and evaluated regularly (Arslan 2017). Natural and anthropogenic factors significantly influence groundwater quality, and such effects may impair groundwater quality for domestic uses and irrigation (Mallick et al. 2018). Irrigation water quality is an essential issue since bad quality water destroys soils and thus plant cover (Barik and Pattanayak 2019).

Irrigation water quality plays an important role in improving soil conditions, crop yields and the development of agricultural economics. Groundwater quality is mainly dependent on anion and cation concentrations. Various classifications are available for irrigation water quality. Electrical conductivity (EC) and sodium adsorption ratio (SAR) are the most significant factors used in these classifications.

✉ Sevda Taşan
sevda.safi@omu.edu.tr

¹ Department of Agricultural Structures and Irrigation,
Faculty of Agriculture, Ondokuz Mayıs University, Samsun,
Turkey

Besides these parameters, residual sodium carbonate (RSC), Kelly index (KI), percentage of sodium (Na%), magnesium ratio (MR), potential salinity (PS) and permeability index (PI) are also used in irrigation water quality assessments and classifications. Many studies have assessed the groundwater quality to investigate its suitability for irrigation (Bhunia et al. 2018; Mokoena et al. 2020; Sarkar et al. 2022). Lanjwani et al. (2020) conducted a study in Pakistan with groundwater samples taken from 25 groundwater wells and investigated water quality parameters (SAR, Na%, KI and PI) to assess the suitability of groundwaters for irrigation. The results of water quality for irrigation indicated that 60–80% of the samples were suitable for irrigation.

It is crucial to investigate the groundwater quality of any region to determine its suitability for drinking, agricultural and industrial purposes (Todd 1959). However, various challenges are encountered during assessing water quality in a region, such as large-scale sample collection and data processing, which is often time-consuming and expensive in terms of equipment, chemicals and human resources (Tiyasha et al. 2020). Therefore, there is an urgent need to develop a model for estimating the water quality parameters of groundwater (Awadh et al. 2020). In this sense, artificial intelligence techniques can be an effective and reliable approach for estimating groundwater quality. Various studies based on the use of artificial intelligence models such as ANN (Al-Waeli et al. 2022; M'nassri et al. 2022), ANFIS (Banadkooki et al. 2020; Elzain et al. 2021) and machine learning (ML) (El Bilali et al. 2021; Yu et al. 2022) in the field of groundwater quality modeling have been conducted by different researchers in parts of the world in recent years. The results of previous research have shown that artificial intelligence (AI) methods are powerful tools for modeling groundwater quality parameters. Still, the models' accuracy depends on the input data used and an appropriate model structure.

Another important requirement in the assessment of groundwater quality is to create a map to identify the areas where groundwater is not suitable for irrigation. These maps are useful tools for the spatial distribution of groundwater quality indices and comparative analysis (Ghazaryan et al. 2020). Geographical information system (GIS) is largely used to gather, analyze and display spatial data and to use these data in decision-making processes in various disciplines, including geology, geo-environment, etc. (Adimalla and Taloor 2020). GIS is a strong technology in the display of distribution models and assessing the behavior of groundwater quality parameters of a basin. The inverse distance weighted (IDW), ordinary kriging (OK) and empirical Bayesian kriging (EBK)-like several deterministic and statistical interpolation methods have been developed to facilitate the estimation of unknown points and to form a continuous dataset for spatial assessment of the place (Hossain et al.

2020). In addition, artificial intelligence-based models have been successfully used to model the spatial and temporal variation of groundwater quality by combining with geographic information systems (Gholami et al. 2022). Sahour et al. (2020) compared statistical (multiple linear regression (MLR)) and machine learning (deep neural networks (DNN), extreme gradient boosting (EGB)) techniques to map the spatial distribution of groundwater salinity on the Caspian coast. The results showed that the combination of field sampling and machine learning techniques is useful for modeling and mapping groundwater salinity.

In order to ensure sustainable water management in agricultural production in the study area, it is necessary to determine the suitability of groundwater for irrigation purposes and to take required precautions in a timely manner. Therefore, the main objectives of this study were i) to evaluate the suitability of groundwater quality for agricultural purposes in the study area with various quality indices such as sodium adsorption ratio (SAR), residual sodium carbonate (RSC), percentage of sodium (Na%), potential salinity (PS), magnesium ratio (MR), permeability index (PI) and Kelly index (KI), (ii) to estimate the irrigation water quality indices using ANN and ANFIS models, (iii) to evaluate the performance of each model using accuracy measures and (iv) to model the spatial distribution of irrigation water quality indices using an integrated GIS approach with artificial intelligence-based techniques. These techniques will assist in the implementation of site-specific remedial measures for sustainable irrigation water management.

Materials and methods

Study area

The study area is located within the boundaries of Bozyazi town of Mersin province on the south of the Central Mediterranean Region between 33°3'–33°7' east longitudes and 36°6'–36°8' north latitudes and covers 6.3 km² of agricultural lands (Fig. 1). A typical Mediterranean climate is dominant in the study area; summers are hot and dry, and winters are warm and rainy. According to long-term averages for Mersin province, December is the rainiest month (139.6 mm) and August is the driest month (6.9 mm). Annual total precipitation is 615.8 mm. The greatest average temperature is recorded in August (31.5 °C) and the lowest in January (6.2 °C) (Anonymous 2020). Besides banana, peanut, citrus species and early vegetable production are common in the region. There are not any irrigation and drainage facilities in the study area, and irrigations are performed only with groundwaters. It is also the first study to investigate the hydrochemical properties of groundwater in this region.

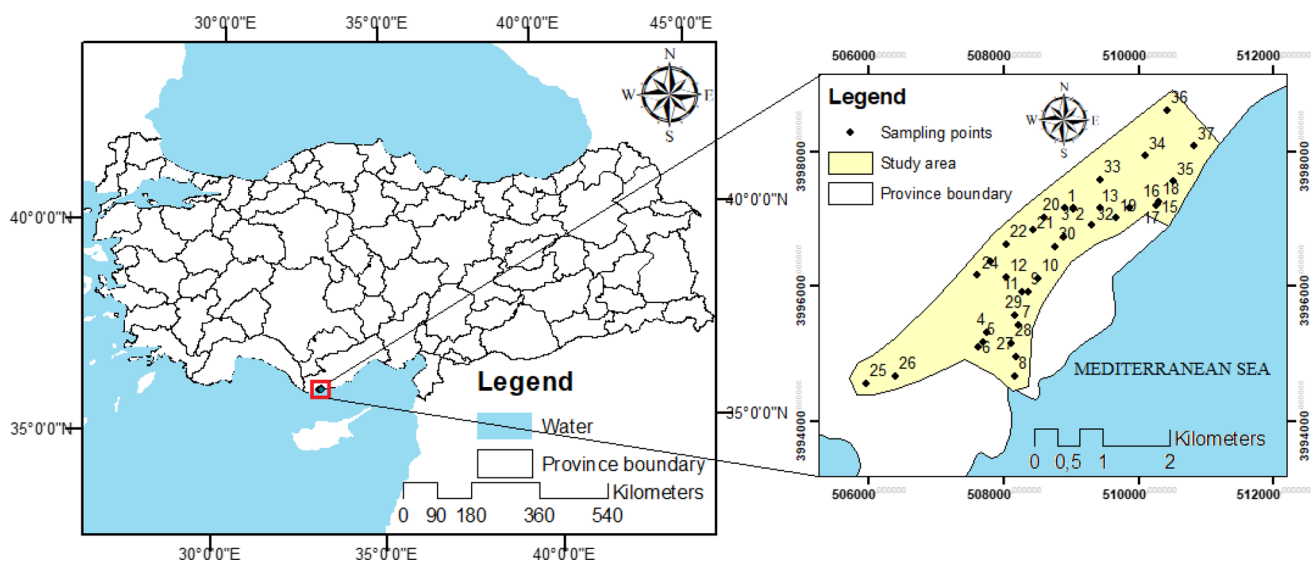


Fig. 1 Location of the studied area

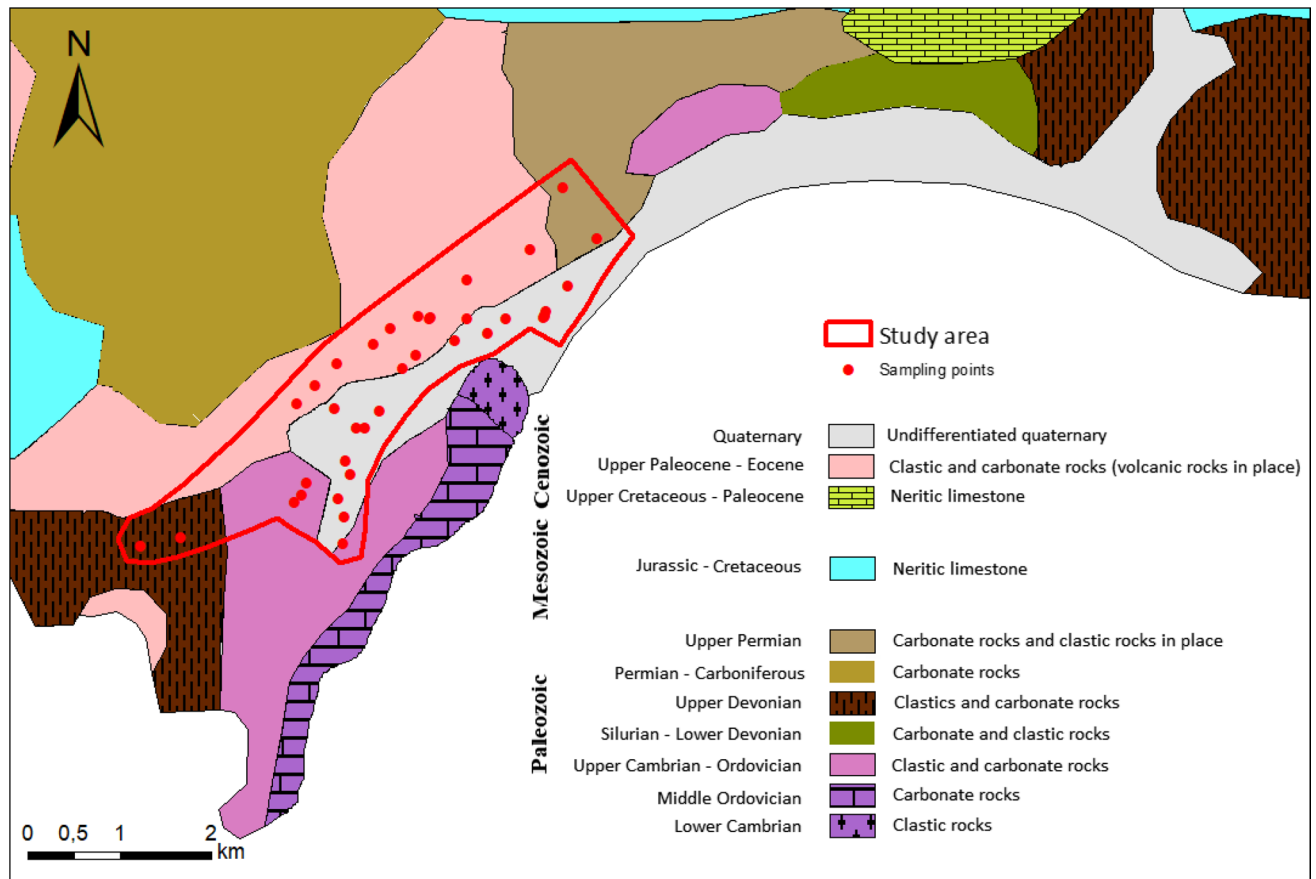


Fig. 2 Geologic map of the study area (modified from geologic map of MTA, 1/500 000 scale)

The geology of the study area is constituted by the rocks ranging in age from Cambrian to Recent or Quaternary deposits (Fig. 2). These rocks consist of Early Paleozoic

carbonates and clastics and platform carbonates from Mesozoic to Early Tertiary. The quaternary formations are represented by undifferentiated quaternary, while the Ordovician

and the Eocene sediments are represented by clastic and carbonate rocks. The aquifer systems are developed in the clastic and carbonate rocks (such as dolomites) and undifferentiated quaternary deposits (Fig. 2).

Sampling and data

Water samples were taken from 37 deep wells before the irrigation season (June 2018). According to the central limit theorem (CLT), as the sample size (n) increases, the sampling distributions will approach the normal distribution. This should be true regardless of whether the source population is normal or skewed, provided the sample size is sufficiently large (usually $n \geq 30$) (Kwak and Kim 2017). Therefore, considering this theorem and the total size of the study area, the number of samples was determined as 37. The locations of all sampling points are illustrated in Fig. 1.

Wells were operated for 10 min, and then, the samples were taken into polyethylene bottles and preserved at +4 °C until the analyses. Before the analyses, the samples were filtered through 0.45- μm acetate cellulose filter papers. To determine the characteristics of groundwaters, the samples were subjected to EC (dS/m), pH, Na^+ (meq/L), K^+ (meq/L), Ca^{2+} (meq/L), Mg^{2+} (meq/L), CO_3^{2-} (meq/L), HCO_3^- (meq/L), SO_4^{2-} (meq/L) and Cl^- (meq/L) analyses. All analyses were performed according to the standard procedures given in the American Public Health Association manual (APHA 2005). Groundwater pH values were measured with a pH meter (Jenway-3510), and EC was measured with a conductivity meter (Jenway-4510). Ca^{2+} and Mg^{2+} concentrations were determined with EDTA titration. Na^+ and K^+ concentrations were determined with the use of a flame photometer. Sulfuric acid titration was used to determine carbonate (CO_3^{2-}) with phenolphthalein indicator and bicarbonate (HCO_3^-) with methyl orange indicator. Sulfate (SO_4^{2-}) content was determined by the UV spectrophotometric method. Chloride (Cl^-) was determined with the use of potassium chromate indicator and silver nitrate titration.

Irrigation water quality parameters

To assess the suitability of groundwaters for irrigation, SAR, RSC, Na%, PS, MR, PI and KI values were calculated using anion and cation concentrations. The following equations were used to calculate the relevant parameters [Eqs. (1)–(7)]:

Richards (1954)

$$\text{SAR} = \frac{\text{Na}^+}{\sqrt{\frac{(\text{Ca}^{2+} + \text{Mg}^{2+})}{2}}} \quad (1)$$

Raghunath (1987)

$$\text{RSC} = [\text{HCO}_3^- + \text{CO}_3^{2-}] - [\text{Ca}^{2+} + \text{Mg}^{2+}] \quad (2)$$

Wilcox (1955)

$$\text{Na}\% = \frac{\text{Na}^+ + \text{K}^+}{\text{Ca}^{2+} + \text{Mg}^{2+} + \text{Na}^+ + \text{K}^+} \times 100 \quad (3)$$

Delgado et al. (2010)

$$\text{PS} = \left[\text{Cl}^- + \frac{1}{2} \text{SO}_4^{2-} \right] \quad (4)$$

Szabolcs and Darab (1964)

$$\text{MR} = \left[\frac{\text{Mg}^{2+}}{\text{Ca}^{2+} + \text{Mg}^{2+}} \right] \quad (5)$$

Doneen (1964)

$$\text{PI} = \left[\frac{\text{Na}^+ + \sqrt{\text{HCO}_3^-}}{\text{Ca}^{2+} + \text{Mg}^{2+} + \text{Na}^+} \right] \quad (6)$$

Kelly (1940)

$$\text{KI} = \left[\frac{\text{Na}^+}{\text{Ca}^{2+} + \text{Mg}^{2+}} \right] \quad (7)$$

In Eqs. (1)–(7), the concentration of all parameters was expressed in meq/L.

Modeling of irrigation water quality parameters

In this study, two artificial intelligence methods are applied for the estimation of irrigation water quality parameters, namely artificial neural networks (ANN) and adaptive neuro-fuzzy inference system (ANFIS). The optimal structure of the models was determined using a trial-and-error procedure. Both methods were tested by comparing the predicted quality parameters values to the measured values in the testing subset using five statistical coefficients. These methods are briefly detailed in the following sections.

Artificial neural network

Feedforward–backpropagation artificial neural network architecture was used in this study to estimate groundwater quality parameters. Multilayer perceptron (MLP) is a common ANN approach used in estimation studies. A multilayer perceptron network comprises an input layer, one or more hidden layers and an output layer. MLP may have more than one hidden layer. Theoretical studies confirm that a single hidden layer is sufficient for the MLP model (Hornik et al. 1989). Therefore, a single hidden layer was used in the present study for MLP. A three-layer

MLP architecture generated within the scope of the present study is presented in Fig. 3. A series of neurons and nodes were arranged in each layer. Neurons of each layer are connected to neurons of the subsequent layer with weights.

The training algorithm is a mathematical formulation optimizing error functions to change connection weights (Dinkar 2017). Although there are many different training algorithms in the literature, it is not always easy to determine which training algorithm will give more accurate results for a particular problem (Haykin 1999; Saraçoğlu 2008). The long training time and the need for many iterations of the gradient descent technique used in most backpropagation applications are the disadvantages of this method. Another backpropagation technique, Levenberg–Marquardt’s (LM) weight optimization approach, offers a solution to overcome this problem (Cigizoglu and Kişi 2005). In the present study, Levenberg–Marquardt backpropagation (LM) training algorithm was used to train ANN. As a transfer function, *tansig* was used in the input layer and *purelin* was used in the output layer. The mathematical function of the ANN is given by:

$$\text{Output} = \sum_{k=1}^m \left(\left[\frac{2}{1 + \exp \left(-2 \left(\sum_{j=1}^m \sum_{i=1}^n ((iw(i,j)x(i) + b_1(j))) \right) \right) - 1 \right] lw(k)b_2 \right) \tag{8}$$

where *iw*, *lw* and *b*₁, *b*₂ are the connection weights and the bias of the networks, respectively; *x* is the number of

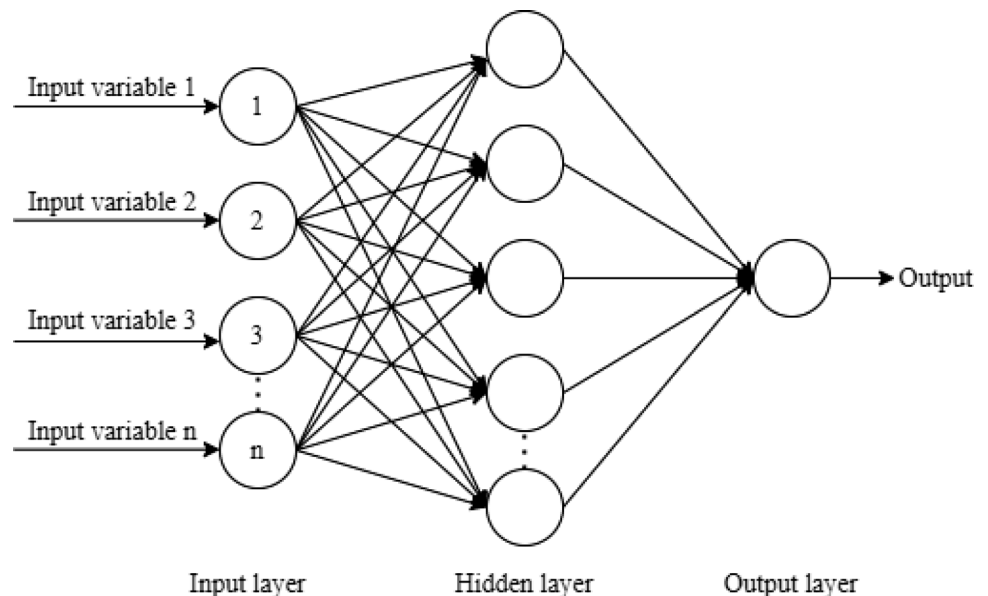
input variables; and *n* and *m* are the number of neurons in the input and hidden layers, respectively. ANN construction parameters (*iw*, *lw*, *b*₁ and *b*₂) obtained for each water quality parameter after adequate data training are given in S-Tables 1–7 (Online Resource 1–7).

The optimum number of neurons and training iterations in the hidden layer is an essential indicator of ANN modeling (Yıldız and Karakuş 2020). In the present study, empirical equations of Wanas et al. (1998) and Mishra and Desai (2006) were used to determine the number of neurons in the hidden layer and trials were conducted between log (*N*) and (2*n* + 1), where *N* is the number of samples in dataset and *n* is the number of input neurons. Then, the optimum number of hidden neurons was selected through the trial-and-error method. Separation of data into training and testing datasets may have significant effects on model outcomes. Therefore, the measured dataset was divided into two subgroups: 70% of the data for training and 30% of the data for testing were used. The training and testing data were divided randomly.

Adaptive neuro-fuzzy inference system (ANFIS)

Adaptive neuro-fuzzy inference system (ANFIS), introduced for the first time by Jang (1993), is a combination of an adaptive neuro-fuzzy inference system. The ANFIS used in

Fig. 3 Three-layer MLP architecture



the present study is a first-order Sugeno fuzzy model (Jang 1993). ANFIS takes advantage of a feedforward network to investigate the better performing fuzzy decision rules. ANFIS uses a specified input–output dataset and forms a FIS in which membership functions are arranged using either a backward propagation algorithm alone or a combination of a backward propagation algorithm and the least-squares method (Abdulshahed et al. 2015). Such a case allows the fuzzy systems to learn from the modeled data. The Sugeno fuzzy structure of the ANFIS model is composed of five layers (Fig. 4). ANFIS is designed with the use of various model definition approaches of the Sugeno model, including grid partition (GP) or subtractive clustering (SC). In the present study, GP method was used. In the ANFIS–GP model, the input section was divided into rectangular subsections using various local fuzzy regions. In the present study, three–four variables were used as input variables to estimate water quality parameters and thus ANFIS-GP could be applied efficiently. Detailed information about ANFIS model could be found in Jang (1993). MATLAB (MathWorks, ver. R2015b) was used for ANN and ANFIS analysis.

Input combinations and data normalization

The ANN and ANFIS methods were used to estimate irrigation water quality parameters (MR, SAR, KI, PS, RSC, Na% and PI). Identifying the input variables is the first stage in creating a prediction model. In this study, suitable input combinations were determined by correlation analysis (Asadollah et al. 2021). AI-based models are highly sensitive to strong correlations between input variables, and principal component analysis was used to solve this potential problem.

Table 1 Input combinations used to model irrigation water quality parameters

Model abbreviation	Variable	Input parameters	
ANN	ANFIS	SAR	EC, pH Na, K
ANN	ANFIS	RSC	EC, K, Mg, Cl
ANN	ANFIS	Na%	pH, Na, HCO ₃ , Cl
ANN	ANFIS	PS	EC, K, Mg, Cl
ANN	ANFIS	MR	pH, Ca, CO ₃
ANN	ANFIS	PI	EC, Na, Cl
ANN	ANFIS	KI	pH, Na, Ca, Cl

Accordingly, for the estimation of each irrigation water quality parameter, physicochemical properties that have a strong correlation with the irrigation water quality indices were determined as input and are summarized in Table 1.

Table 1. Input combinations used to model irrigation water quality parameters.

The normalization was done for all datasets to improve the modeling performance. The normalization equation is defined as follows:

$$X_n = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} \tag{9}$$

where X_i is the observed value, X_n is the normalized value of X_i and X_{\max} and X_{\min} are the maximum and minimum values, respectively.

Figure 5 shows the flow chart of the artificial intelligence models used.

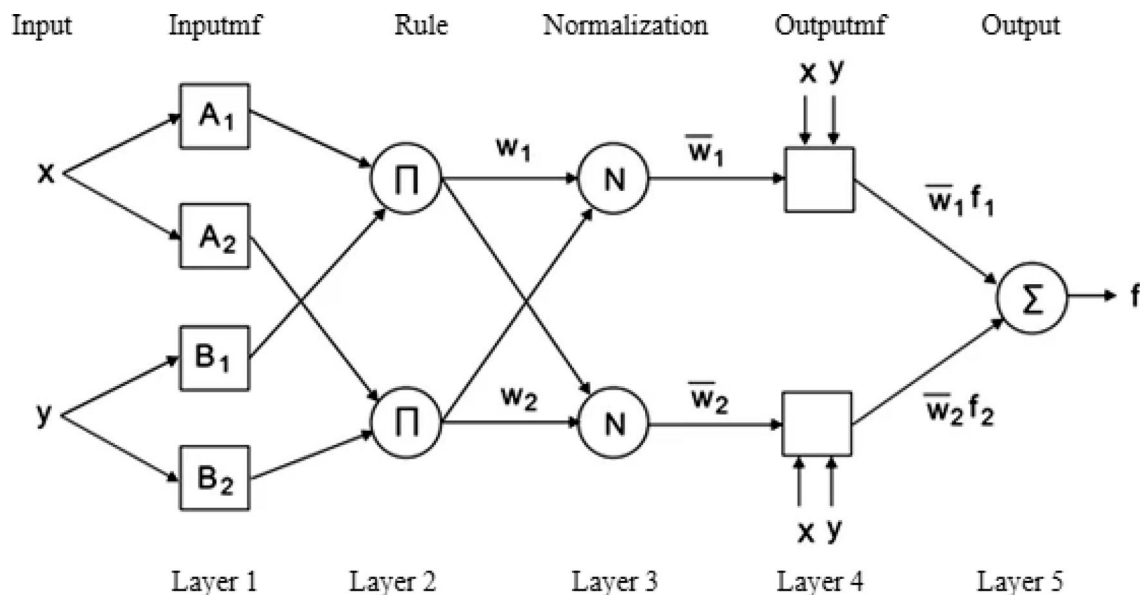
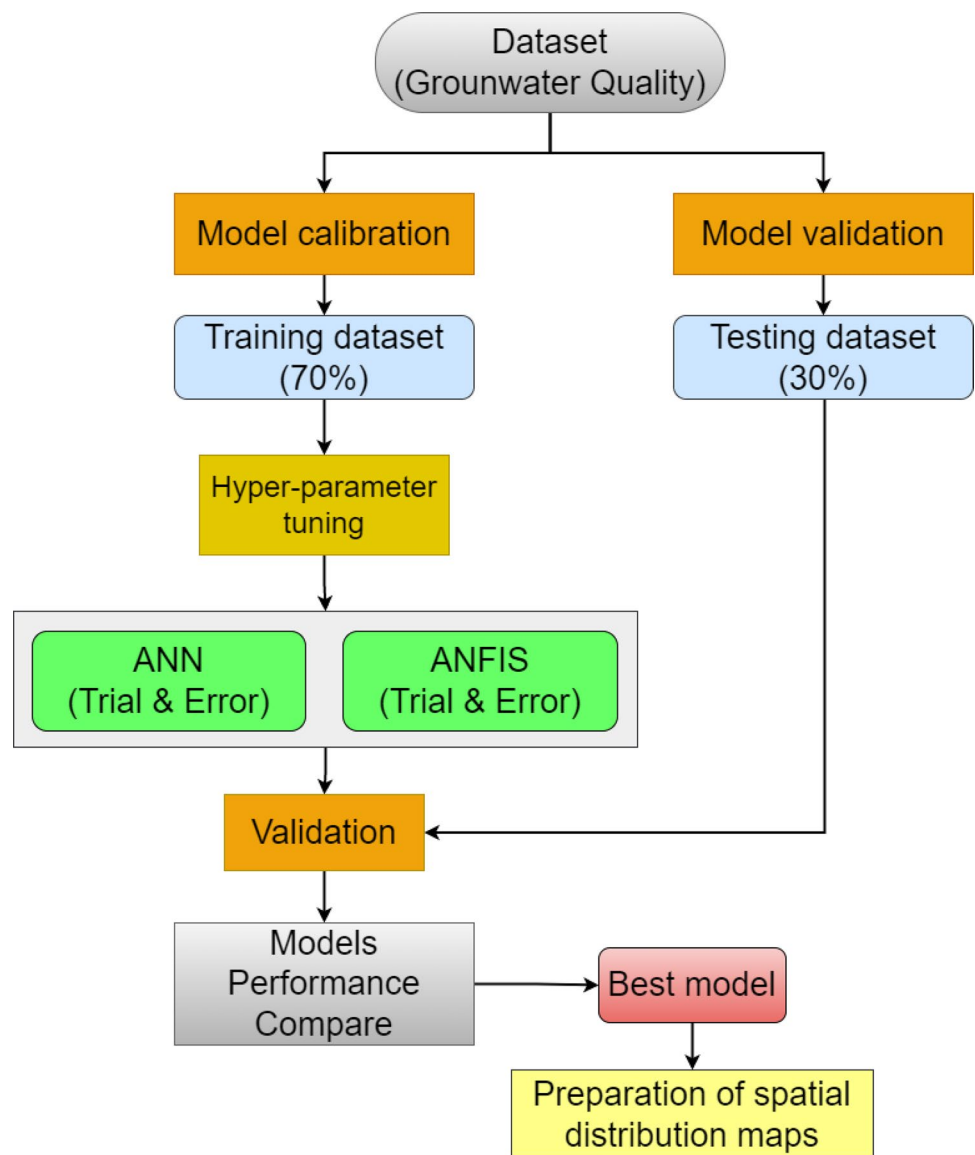


Fig. 4 Adaptive neuro-fuzzy inference system (ANFIS) structure

Fig. 5 Flowchart for artificial intelligence methods and selection of the best model for estimation of groundwater quality parameters



Spatial distribution maps

LI and Heap (2011) reviewed the studies comparing different interpolation methods in environmental sciences and indicated inverse distance weight (IDW), ordinary kriging (OK) and ordinary co-kriging (OCK) methods as the most common methods used. In the present study, ArcGIS 10.2 software and the IDW method were used to generate spatial distribution maps of measured and estimated irrigation water quality parameters. This method assumes that the value of an unsampled point was the weighted average of the values of sampled points surrounding that unsampled point (neighborhood) (Longley et al. 2010). Estimated values are a function of the distance and size of neighboring points, and the influence on estimation points decreases

with increasing distances. For IDW, the power function was selected as 2, which is the most common and assumed in ArcGIS, and interpolation maps were generated using a maximum 15 neighboring points.

Performance criteria

Five different statistical parameters were used to assess the performance of ANN and ANFIS models: coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE) and Nash–Sutcliffe efficiency coefficient (NS). Coefficient of determination [Eq. (10)] defines the correlation degrees between the estimated and observed values (Jahani et al. 2016):

$$R^2 = \left(\frac{\sum_{i=1}^N (O_i - O)(P_i - P)}{\sqrt{\sum_{i=1}^N (O_i - O)^2 \sum_{i=1}^N (P_i - P)^2}} \right)^2 \tag{10}$$

Root mean squared error (RMSE) (Chai and Draxler 2014) can be calculated as follows [Eq. (11)]:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (O_i - P_i)^2}{N}} \tag{11}$$

The mean absolute error (MAE) (Chai and Draxler 2014) can be calculated as follows [Eq. (12)]:

$$MAE = \frac{\sum |O_i - P_i|}{N} \tag{12}$$

Nash–Sutcliffe efficiency coefficient (NS) (Nash and Sutcliffe 1970) can be calculated as follows [Eq. (13)]:

$$NS = 1 - \frac{\sum_{i=1}^N (O_i - P_i)^2}{\sum_{i=1}^N (O_i - O)^2} \tag{13}$$

where O_i is the observed value, P_i is the estimated value, O is the mean of observed value, P is the mean of estimated value, N is the number of observations and k is the number

of free parameters used in those models. The best compliance between the estimated and observed values is achieved at $R^2 = 1$, $RMSE = 0$, $MAE = 0$ and $NS = 1$.

Results and discussion

Physical and chemical characteristics of irrigation water

Physical and chemical characteristics of groundwaters and minimum, maximum, mean and standard deviation of irrigation water quality parameters are provided in Table 2 for training and testing datasets. Mean values of randomly selected training and testing data were close to each other.

Electrical conductivity (EC) values at 25 °C ranged from 433 to 1975 dS/m with an average value of 938.54 dS/m for the training dataset and 842.18 dS/m for the testing dataset. The pH is an important indicator of water nature and is closely related to the other chemical parameters (Adimalla and Wu 2019). The pH of groundwater samples varied between 7.23 and 7.92.

Table 2 Descriptive statistics of input and output parameters

Groundwater characteristics	Training				Testing			
	Min ^a	Max ^b	Mean	SD ^c	Min ^a	Max ^b	Mean	SD ^c
<i>Input</i>								
EC (µS/cm)	433.00	1975.00	938.54	280.27	435.00	1173.00	842.18	215.20
pH	7.23	7.92	7.65	0.15	7.44	7.84	7.67	0.14
Na ⁺ (meq/L)	0.74	8.74	1.76	1.50	0.87	1.96	1.26	0.35
K ⁺ (meq/L)	0.01	0.15	0.04	0.03	0.01	0.06	0.03	0.01
Ca ²⁺ (meq/L)	1.57	7.06	4.36	1.32	1.72	5.43	3.62	1.09
Mg ²⁺ (meq/L)	2.02	10.00	5.93	1.71	5.09	7.87	6.52	0.99
CO ₃ ⁻ (meq/L)	0.00	1.92	0.73	0.47	0.20	1.44	0.75	0.38
HCO ₃ ⁻ (meq/L)	2.00	6.64	4.62	1.26	1.68	6.96	4.07	1.25
Cl ⁻ (meq/L)	1.83	15.55	6.62	2.40	4.27	7.72	6.30	1.08
SO ₄ ²⁻ (meq/L)	0.03	0.55	0.12	0.12	0.02	2.09	0.31	0.59
<i>Output</i>								
MR	39.19	79.37	57.47	9.80	52.83	74.69	64.82	6.98
SAR	0.33	3.40	0.79	0.60	0.38	0.82	0.56	0.15
KI	0.06	0.66	0.18	0.14	0.08	0.22	0.13	0.04
PS	1.88	15.62	6.68	2.40	4.35	7.82	6.46	0.98
RSC	- 7.87	- 0.96	- 4.95	1.58	- 6.44	- 3.43	- 5.32	0.89
Na%	5.88	40.27	14.71	8.08	7.99	18.34	11.37	3.11
PI	21.97	55.21	32.68	7.95	24.69	33.83	28.69	2.73

^aMin minimum value

^bMax maximum value

^cSD standard deviation

Hydrochemical classification

Piper trilinear diagrams (Piper 1944) have been commonly used to understand the hydrogeochemical regimes of research areas (Wu et al. 2018). The present study used the Piper diagram to determine the hydrochemical types of groundwater in the study area. As shown by the Piper diagram (Fig. 6), in all groundwater samples taken from the study area, the anions were dominated by Cl^- and the cations were dominated by Mg^{2+} . Chloride is found in almost all waters and is obtained from a number of sources, including natural mineral deposits, seawater intrusion, agricultural or irrigation discharges and industrial effluents (Malcolm et al. 2007). In the studies, it was concluded that the groundwater might be under the influence of seawater in cases where the Cl amount is above 250 mg/L (Andreassen et al. 1997). The major source of Mg^{2+} in groundwater is due to ion exchange of minerals in rocks and soils by water (Boateng et al. 2016). The majority of the samples fell in mixed Ca–Mg–Cl which indicated ion exchange and also reverse ion exchange, interactions of rock–water, unsaturated zones reactions and anthropogenic influences.

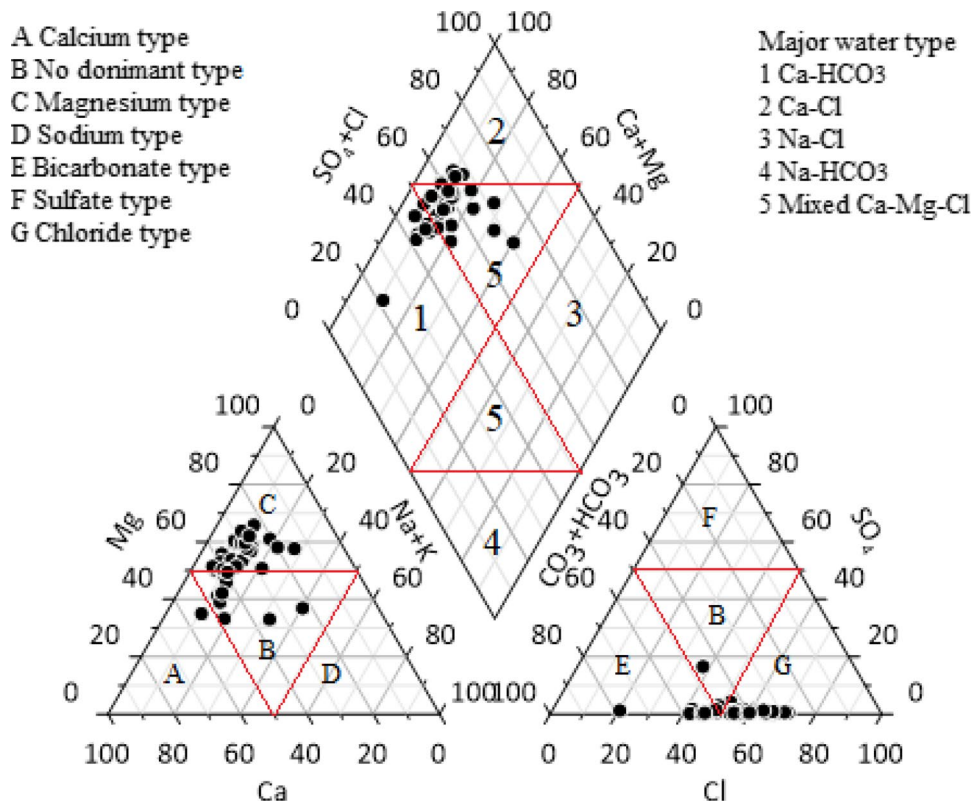
Processes influencing groundwater chemistry

Gibbs diagrams are widely used to demonstrate the relationships between the water composition and the lithological features of the aquifer (Gibbs 1970). In general, precipitation, evaporation and rock weathering are the three major natural mechanisms determining the water chemistry in the study area (Fig. 7). Gibbs plot shows that the rock–water interaction and secondly evaporation largely control groundwater in the study area. This indicates that the dissolution of minerals is the most important process controlling groundwater hydrogeochemistry in the study area.

Correlation analysis

One of the most important steps in developing a data-based model is the selection of significant model input variables (Muttill and Chau 2007). Correlations analysis (Pearson or spearman) could be conducted to select the input parameters used in such models. The correlation coefficient is a numerical value indicating the statistical relationships between two or more variables (Rauf et al. 2018). In the present study, such correlations were investigated between 10 input and 7 output parameters. Data normality was checked with the use of the Kolmogorov–Smirnov test. The present data did not have a normal distribution; therefore, Spearman’s correlation analysis was used. The

Fig. 6 Piper diagram of groundwater samples



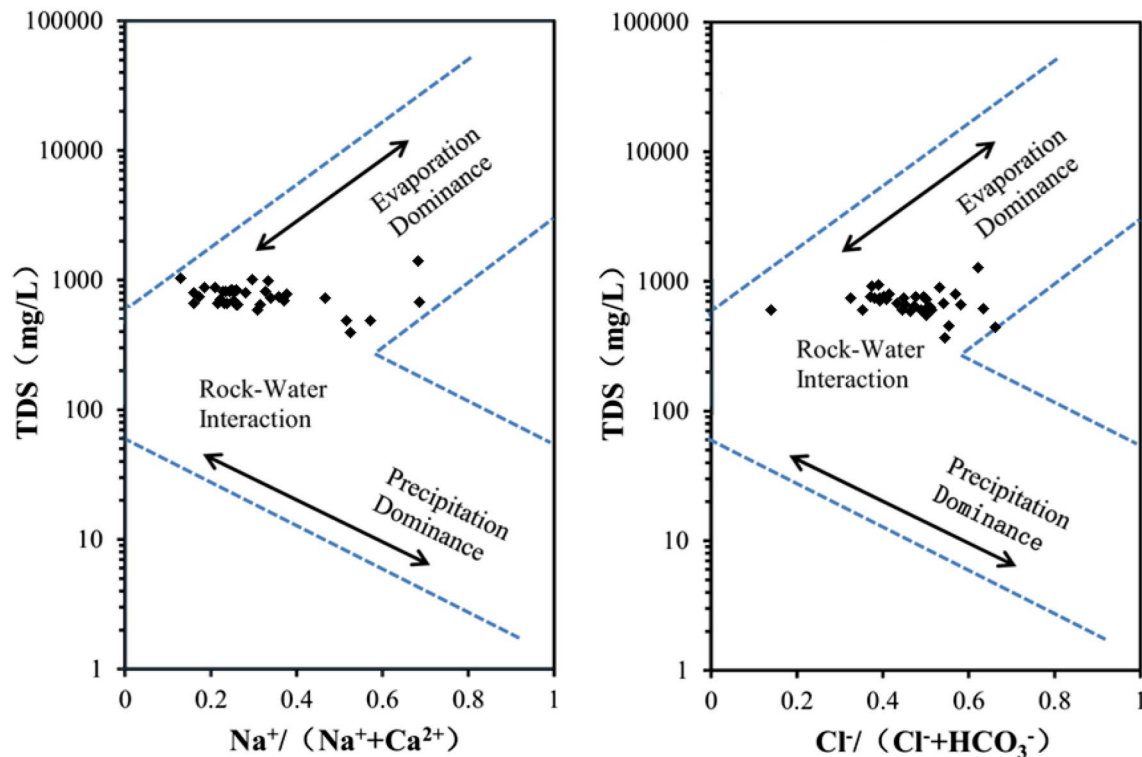


Fig. 7 Gibbs diagram for groundwater samples

Table 3 Spearman correlation coefficients of water quality parameters

	EC	pH	Na ⁺	K ⁺	Ca ²⁺	Mg ²⁺	CO ₃ ²⁻	HCO ₃ ⁻	Cl ⁻	SO ₄ ²⁻
MR	-0.417*	0.456**	0.107	0.288	-0.720**	0.536**	0.060	-0.315	0.243	-0.060
SAR	-0.097	-0.078	0.967**	0.184	-0.219	-0.198	-0.111	-0.247	0.070	0.099
KI	-0.232	-0.086	0.899**	0.134	-0.306	-0.294	-0.178	-0.344*	-0.026	0.093
PS	0.328*	0.111	0.195	0.485**	0.256	0.694**	-0.114	0.146	0.988**	0.096
RSC	-0.365*	-0.100	0.111	-0.313	-0.298	-0.738**	0.042	-0.165	-0.869**	-0.099
Na%	-0.235	-0.078	0.900**	0.147	-0.306	-0.292	-0.180	-0.347*	-0.020	0.093
PI	-0.289	-0.208	0.656**	-0.019	-0.304	-0.565**	-0.237	-0.211	-0.365*	0.059

*Significant at $p < 0.05$

**Significant at $p < 0.01$

correlations between irrigation water quality parameters and physicochemical characteristics of groundwater are provided in Table 3. MR had negative correlations with EC (-0.417) and Ca²⁺ (-0.720) and positive correlations with pH (0.456) and Mg²⁺ (0.536). SAR had highly significant positive correlations with Na⁺ (0.967) ($p < 0.01$). KI had positive correlations with Na⁺ (0.899) and negative correlations with HCO₃⁻ (-0.344). PS had highly significant positive correlations with K⁺ (0.485), Mg²⁺ (0.694) and Cl⁻ (0.988) ($p < 0.01$) and significant positive correlations with EC (0.328) ($p < 0.05$). RSC had negative correlations with EC (-0.365), Mg²⁺ (-0.738) and

Cl⁻ (-0.869). Na% had highly significant positive correlations with Na⁺ (0.900) ($p < 0.01$) and significant negative correlations with HCO₃⁻ (-0.347) ($p < 0.05$). PI had highly significant positive correlations with Na⁺ (0.656) ($p < 0.01$) and negative correlations with Mg²⁺ (-0.565) and Cl⁻ (-0.365). According to Table 2, the parameters with the greatest effect on irrigation water quality indices were identified as Na⁺, EC, Cl⁻, Mg²⁺, Ca²⁺, HCO₃⁻ and pH and the parameters with the least effect were identified as K, CO₃ and SO₄.

Principal component analysis

Factor analysis has been successfully used to obtain significant correlations among hydrochemical characteristics of groundwater samples (Chen et al. 2006; Mondal et al. 2010). Jahin et al. (2020) applied multivariate analysis to improve the irrigation water quality index (IWQI) in Egypt. The results have shown that using principal component analysis (PCA) and factor analysis (FA) can control water quality quickly and inexpensively. In the present study, principal component analysis was used to determine the input parameters of the artificial intelligence-based models to be generated for the estimation of irrigation water quality indices. Physical and chemical characteristics of EC, pH, Na⁺, K⁺, Ca²⁺, Mg²⁺, SO₄²⁻, CO₃⁻, HCO₃⁻ and Cl⁻ and irrigation water quality indices of MR, SAR, KI, PS, RSC, Na% and PI were used in the principal component analysis. The results of principal component analysis conducted with 37 groundwater samples and 17 parameters are provided in Table 3. There were five factors with eigenvalue of greater than 1. These five factors explained 88.13% of the total variation. The first eigenvalue explaining 32.83% of total variance constituted the first and the main factor. The second, third, fourth and fifth eigenvalues were calculated as 4.71, 2.96, 1.57 and 1.27

Table 4 Rotated factor structure matrix

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Na%	0.93*	0.13	0.24	- 0.19	0.01
KI	0.93*	0.20	0.20	- 0.19	0.03
PI	0.90*	- 0.21	0.19	- 0.25	- 0.01
SAR	0.89*	0.42	0.08	- 0.09	0.02
Na ⁺	0.80*	0.56	- 0.03	- 0.01	0.02
PS	0.28	0.95*	- 0.08	- 0.01	- 0.03
Cl ⁻	0.29	0.94*	- 0.08	- 0.02	- 0.10
RSC	0.35	- 0.87*	0.09	- 0.01	- 0.05
Mg ²⁺	- 0.27	0.76*	0.07	0.53	0.01
K ⁺	0.59	0.63**	0.10	0.10	- 0.06
EC	0.26	0.63**	- 0.60	0.23	0.21
Ca ²⁺	- 0.19	0.23	- 0.94*	0.05	0.06
MR	- 0.04	0.32	0.83*	0.34	- 0.06
HCO ₃ ⁻	- 0.11	0.18	- 0.74*	0.39	- 0.02
CO ₃ ⁻	- 0.06	0.05	- 0.07	0.84*	0.10
SO ₄ ²⁻	- 0.12	0.00	0.01	0.15	0.93*
pH	- 0.46	0.08	0.35	0.11	- 0.56
Eigenvalue	6.24	4.71	2.96	1.57	1.27
% Total variance	32.83	24.81	15.57	8.27	6.66
Cumulative % variance	32.83	57.64	73.21	81.48	88.13

*Strong ($p > 0.75$)

**Medium (0.50–0.75)

and explained 24.81, 15.57, 8.27 and 6.66% of the total variation, respectively. The magnitude of eigenvalues was used as a criterion to interpret the relationship between the groundwater quality parameters and the factors. Liu et al. (2003) classified factor loads as: “strong” for > 0.75, “medium” for 0.50–0.75 and “poor” for 0.30–0.50. The input parameters to be used in the estimation of each quality index were determined with the use of correlation analysis results and factor loads, as indicated with bold values in Table 4.

Artificial neural network (ANN) models for groundwater quality

The input parameters used in estimating quality parameters to assess the suitability of groundwaters for irrigation and the best number of neurons used in the hidden layer and the number of iterations are summarized in Table 5. The input parameters used in the estimation of irrigation water quality indices were selected based on the results of principal component analysis and correlation analysis. The number of neurons in the hidden layer was determined with the use of the trial-and-error method. The number of iterations was tried to be between 20 and 50 in 10 increments, and resultant outcomes were compared to get the best number of iterations. The parameters used in selecting the optimum number of neurons in the hidden layer and the best number of iterations have the greatest R² and the lowest RMSE and MAE values. The performance criteria calculated for each quality parameter in the testing dataset are presented in Fig. 8. The performance parameters yielding the best outcomes for the ANN model are summarized in Table 5.

For estimation of SAR, the greatest R² and the lowest RMSE and MAE values were achieved with the model with five neurons in the hidden layer and 50 iterations (Fig. 8a). The model with the best network architecture of 4–5–1 had R², RMSE, MAE and NS values of 0.97, 0.05, 0.04 and 0.97, respectively (Table 6). For RSC with four input parameters, five–nine hidden neurons were tried and the best

Table 5 Input parameters used in estimation model, number of hidden neurons and number of iterations

Variable	Input parameters	Best number of hidden nodes	Iteration
SAR	EC, pH, Na ⁺ , K ⁺	5	50
RSC	EC, K, Mg ²⁺ , Cl ⁻	5	50
Na%	pH, Na ⁺ , HCO ₃ ⁻ , Cl ⁻	9	40
PS	EC, K ⁺ , Mg ²⁺ , Cl ⁻	7	30
MR	pH, Ca ²⁺ , CO ₃ ⁻	4	20
PI	EC, Na ⁺ , Cl ⁻	6	30
KI	pH, Na ⁺ , Ca ²⁺ , Cl ⁻	6	50

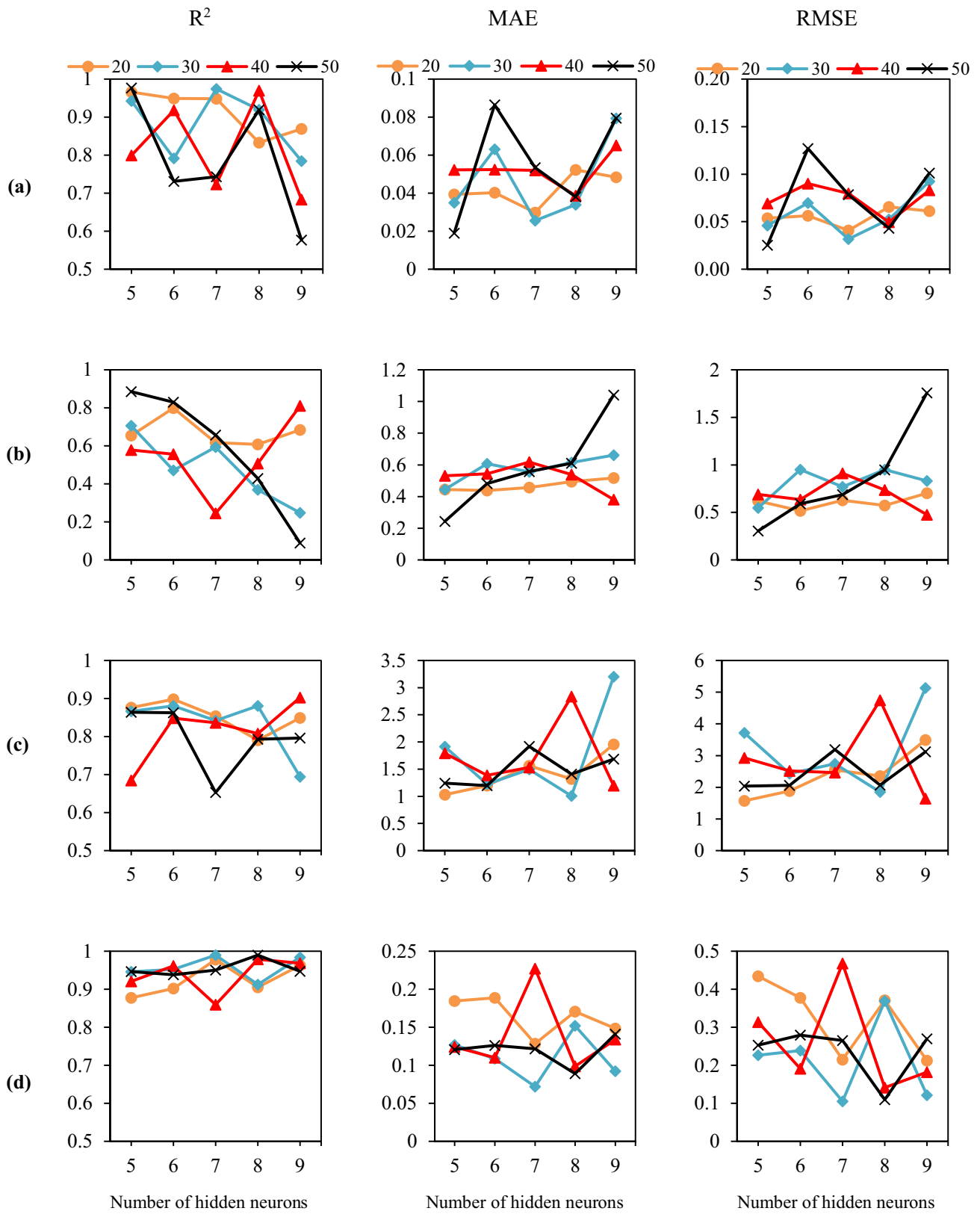


Fig. 8 Effects of number of hidden neurons and iterations in testing dataset of developed ANN models on performance criteria: **a** SAR, **b** RSC, **c** Na%, **d** PS, **e** MR, **f** PI and **g** KI

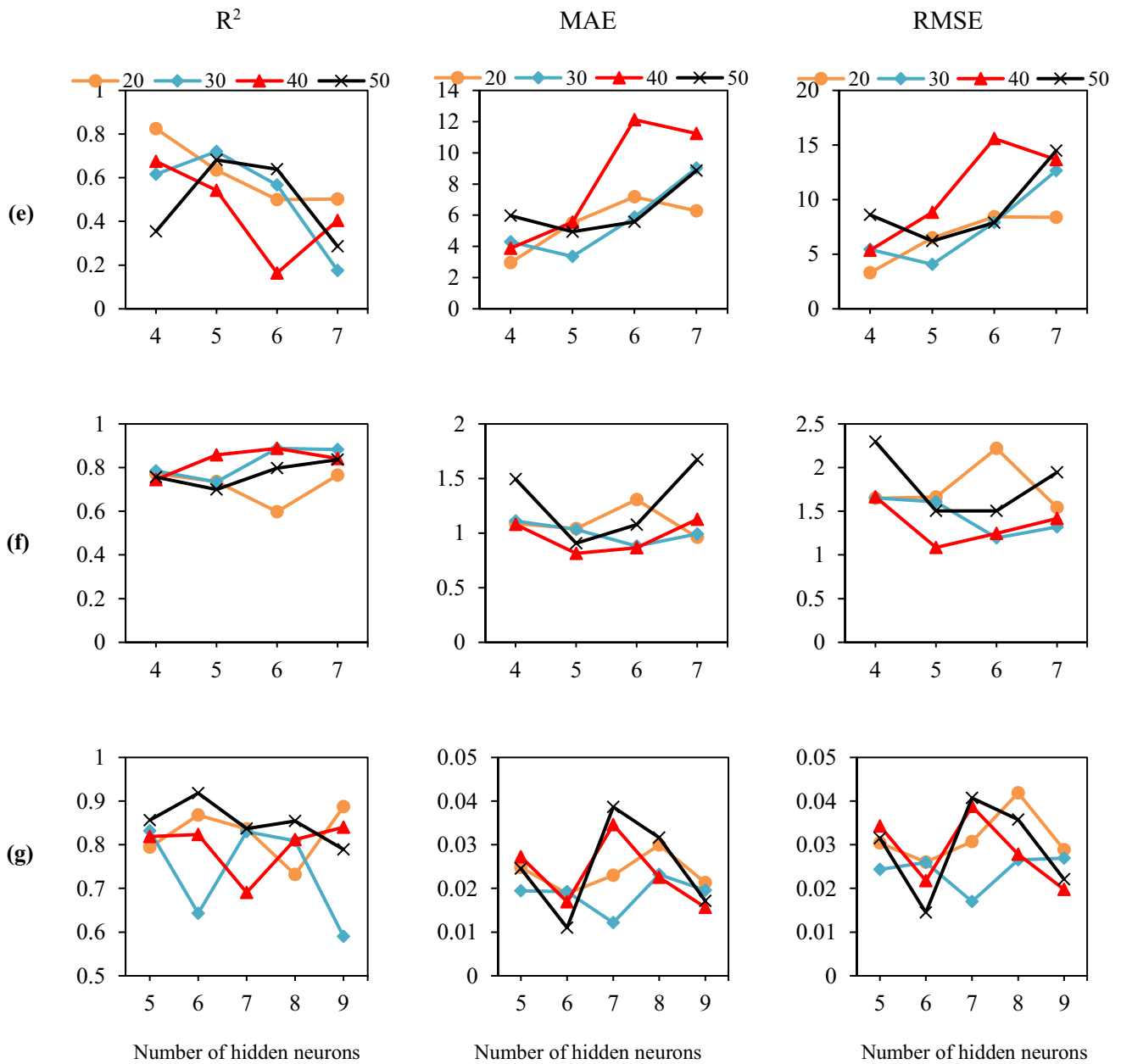


Fig. 8 (continued)

Table 6 Performance criteria for ANN models

Parameters	Network structure	Training				Testing			
		R^2	MAE	RMSE	NS	R^2	MAE	RMSE	NS
SAR	(4–5–1)	1.00	0.01	0.02	1.00	0.98	0.02	0.03	0.97
RSC	(4–5–1)	0.97	0.17	0.25	0.97	0.88	0.24	0.3	0.88
Na%	(4–9–1)	1.00	0.04	0.07	1.00	0.90	1.20	1.64	0.69
PS	(4–7–1)	1.00	0.01	0.02	1.00	0.99	0.07	0.11	0.99
MR	(3–4–1)	0.71	3.80	5.14	0.71	0.83	2.95	3.28	0.76
PI	(3–6–1)	0.99	0.46	0.67	0.99	0.89	0.88	1.19	0.79
KI	(4–6–1)	1.00	0.01	0.01	1.00	0.92	0.01	0.02	0.86

performance criteria were achieved with five hidden neurons and 50 iterations (Fig. 8b). Then, for RCS estimated in the testing dataset, R^2 , RMSE, MAE and NS values were calculated as 0.88, 0.30, 0.24 and 0.88, respectively (Table 6). For the models developed for estimation of Na%, the greatest R^2 was calculated as 0.90 and the lowest RMSE and MAE values were calculated as 1.64 and 1.20, respectively, and such performance criteria were achieved with nine hidden neurons and 40 iterations (Table 6, Fig. 8c). According to the four-input model developed for estimation of PS parameter, the model with the best performance criteria had the network structure of 4–7–1 and such a best performance was achieved with 30 iterations (Fig. 8d). For PS model, R^2 was 0.99, NS was 0.99, RMSE was 0.11 and MAE was 0.07 (Table 6). For the model used in the estimation of MR in the testing dataset, R^2 , RMSE and MAE values were calculated as 0.83, 3.28 and 2.95, respectively, and such criteria were achieved with 3–4–1 network structure and 50 iterations (Fig. 8e). The model used in the estimation of PI trait in which EC, Na^+ and Cl^- were used as input parameters had the best network architecture of 3–6–1 and the best performance was achieved with 30 iterations (Fig. 8f). In this model, R^2 was 0.86, NS was 0.83, RMSE was 1.08 and MAE was 0.81 (Table 6). The model developed for estimation of KI, in which four input parameters were used, had the best network structure of 4–6–1 and the best performance was achieved with 50 iterations (Fig. 8g). In this KI model, R^2 , RMSE and MAE values were calculated as 0.92, 0.02 and 0.01, respectively (Table 6). It was observed that a different number of hidden neurons were effective in the estimation of various parameters. The models developed had a high coefficient of determination and thus yielded more accurate estimations. Figure 8 shows the accuracy and reliability of

the present models developed for parameter estimation. El Bilali and Taleb (2020) estimated irrigation water quality via using exchangeable sodium percentage (ESP), percentage of sodium (Na%), residual sodium carbonate (RSC), permeability index (PI), Kelly index (KI), chloride (Cl^-), magnesium ratio (MR) and total dissolved solids (TDS) parameters in 300 samples from taken nine stations on four main rivers using eight different machine learning methods. EC and pH were used as inputs in models. The artificial neural network model has performed much higher compared with the support vector machine.

Adaptive neuro-fuzzy inference system (ANFIS) models for groundwater quality

Model structures of ANFIS-GP models developed for estimation of irrigation water quality indices with ANFIS with the different number of membership functions and Gaussian membership function types are compared in Table 7. The models developed for SAR estimation yielded better performance with 3*3 number of membership functions. In the testing dataset of this model, R^2 , RMSE and NS values were calculated as 0.65, 0.12 and 0.29, respectively. In the model developed for RSC estimation, better performance was achieved with 3*3 number of membership functions and R^2 , RMSE and NS values of this model were calculated as 0.16, 0.77 and – 0.81, respectively. In the model developed for the estimation of Na%, better performance was achieved with 3*3 number of membership functions and the performance criteria of this model were measured as: $R^2=0.63$, RMSE=1.83 and NS=0.33 (Table 7). The model with the best performance for PS parameter had 3*3 number of membership functions and R^2 , RMSE and NS

Table 7 Performance criteria for ANFIS-GP models

Parameters	Number of membership functions	Membership function type	Training				Testing			
			R^2	MAE	RMSE	NS	R^2	MAE	RMSE	NS
SAR	3*3	gaussmf	1.00	0.00	0.00	1.00	0.65	0.04	0.12	0.29
	4*4	gaussmf	1.00	0.00	0.00	1.00	0.63	0.07	0.09	0.60
RSC	3*3	gaussmf	1.00	0.00	0.00	1.00	0.16	1.14	0.77	– 0.81
	4*4	gaussmf	1.00	0.00	0.00	1.00	0.08	1.77	0.99	– 3.35
Na%	3*3	gaussmf	1.00	0.00	0.00	1.00	0.63	1.83	2.43	0.33
	4*4	gaussmf	1.00	0.00	0.00	1.00	0.78	2.40	3.14	– 0.12
PS	3*3	gaussmf	1.00	0.00	0.00	1.00	0.90	0.22	0.3	0.90
	4*4	gaussmf	1.00	0.00	0.00	1.00	0.86	0.33	0.49	0.73
MR	3*3	gaussmf	1.00	0.00	0.00	1.00	0.54	9.74	13.64	– 3.20
	4*4	gaussmf	1.00	0.00	0.00	1.00	0.01	10.84	14.54	– 3.77
PI	3*3	gaussmf	1.00	0.52	0.32	1.00	0.52	2.40	1.94	0.15
	4*4	gaussmf	1.00	0.29	0.15	1.00	0.09	13.25	5.74	– 24.94
KI	3*3	gaussmf	1.00	0.00	0.00	1.00	0.01	0.06	0.08	– 2.83
	4*4	gaussmf	1.00	0.00	0.00	1.00	0.65	0.05	0.06	– 1.75

values of this model were calculated as 0.90, 0.30 and 0.90, respectively. For MR variables, the best performance was achieved with 3*3 number of membership functions and R^2 , RMSE and NS values of this model in the testing dataset were calculated as 0.54, 13.64 and -3.20, respectively. For the models developed to estimate PI parameter, more accurate estimates were achieved with 3*3 number of membership functions and R^2 , RMSE and NS performance criteria of this model were calculated as 0.52, 1.94 and 0.15, respectively. Finally, for the models developed for estimation of KI variable, better performance was achieved with 4*4 number of membership functions and the performance criteria of this model were calculated as: $R^2=0.65$, $RMSE=0.06$, and $NS=-1.75$ (Table 7).

Comparison of ANN and ANFIS models

Groundwater quality indices were estimated using ANN and ANFIS techniques, and the performance of these methods is compared in Table 8. In the testing stage, the R^2 values of ANN and ANFIS models for SAR parameters were calculated as 0.98 and 0.65 and RMSE values as 0.03 and 0.12, respectively. For RSC parameter in the testing stage, R^2 values of ANN and ANFIS models were measured as 0.88 and 0.16 and RMSE values as 0.30 and 0.77, respectively (Table 8). For Na% in the testing stage, the ANN model had an R^2 value of 0.90 and the ANFIS model had an R^2 of 0.63 (Table 8). For PS parameter in the testing stage, the ANN model had a greater R^2 (0.99) than the ANFIS model ($R^2=0.90$); the ANN model had a lower RMSE (0.11) value than the ANFIS (0.30) model (Table 8). For the MR parameter in the testing stage, the ANN model had an R^2 of 0.83 and RMSE of 3.28 and the ANFIS model had an R^2 of 0.54

and RMSE of 13.64. For PI parameter, the greatest R^2 (0.89) and the lowest RMSE (1.19) were obtained from the ANN model. For KI in the testing stage, R^2 values of ANN and ANFIS models were calculated as 0.92 and 0.65, respectively, and RMSE values were calculated as 0.02 and 0.06, respectively. In the testing stage, ANN models had lower RMSE values than the ANFIS models, and such cases indicated that ANN models had better calibration capacities for relevant data than the ANFIS models. Our findings agreed with the results of Singh et al. (2021). MR is the worst predictive parameter. This may be related to the significant correlation between the input and output variables (Mokhtar et al. 2022). The more significant the correlation between the input and output variables, the higher the performance of the models (Trabelsi and Ali 2022). El Bilali (2021) reported that ANN models are less sensitive to input variables. As a result, it is important to investigate artificial smart models for irrigation water quality index estimation using only physicochemical parameters as input variables.

A Nash–Sutcliffe efficiency coefficient (NS) value of 1 indicates perfect estimations of that model. Normally, NS of greater than 0.8 indicates that the model yielded accurate estimations (Shu and Ouarda 2008). The NS values of ANN and ANFIS models in the testing stage were, respectively, calculated as 0.97 and 0.29 for SAR, as 0.88 and -0.81 for RSC, as 0.69 and 0.33 for Na %, as 0.99 and 0.90 for PS, as 0.76 and -3.20 for MR, as 0.79 and 0.15 for PI and as 0.86 and -1.75 for KI (Table 8). In the testing stage, the majority of NS values of ANN models were greater than 0.8 and such a case indicated that ANN models yielded reliable estimations. The ANN models used to estimate water quality parameters yielded greater NS values than the ANFIS models, and such a case indicated that ANN models had better

Table 8 Comparison of ANN and ANFIS models in training and testing stages

Parameters	Methods	Training				Testing			
		R^2	MAE	RMSE	NS	R^2	MAE	RMSE	NS
SAR	ANN	1.00	0.01	0.02	1.00	0.98	0.02	0.03	0.97
	ANFIS	1.00	0.00	0.00	1.00	0.65	0.04	0.12	0.29
RSC	ANN	0.97	0.17	0.25	0.97	0.88	0.24	0.30	0.88
	ANFIS	1.00	0.00	0.00	1.00	0.16	1.14	0.77	-0.81
Na%	ANN	1.00	0.04	0.07	1.00	0.90	1.20	1.64	0.69
	ANFIS	1.00	0.00	0.00	1.00	0.63	1.83	2.43	0.33
PS	ANN	1.00	0.01	0.02	1.00	0.99	0.07	0.11	0.99
	ANFIS	1.00	0.00	0.00	1.00	0.90	0.22	0.30	0.90
MR	ANN	0.71	3.80	5.14	0.71	0.83	2.95	3.28	0.76
	ANFIS	1.00	0.00	0.00	1.00	0.54	9.74	13.64	-3.20
PI	ANN	0.99	0.46	0.67	0.99	0.89	0.88	1.19	0.79
	ANFIS	1.00	0.52	0.32	1.00	0.52	2.40	1.94	0.15
KI	ANN	1.00	0.01	0.01	1.00	0.92	0.01	0.02	0.86
	ANFIS	1.00	0.00	0.00	1.00	0.65	0.05	0.06	-1.75

Best statistical indicators among all models are marked in bold

general estimation quality than the ANFIS models. Artificial intelligence models such as artificial neural networks (ANN) and adaptive neuro-fuzzy inference (ANFIS) systems can model complex and nonlinear variables sufficiently

accurately (Adamowski et al. 2012). Numerous studies conducted in various disciplines have shown that ANN can produce predictions with higher accuracy than traditional statistical techniques (Sreekanth et al. 2011; Abyaneh et al.

Fig. 9 Scatterplot of measured versus predicted values for training and testing datasets for **a** SAR, **b** RSC, **c** Na%, **d** PS, **e** MR, **f** PI and **g** KI

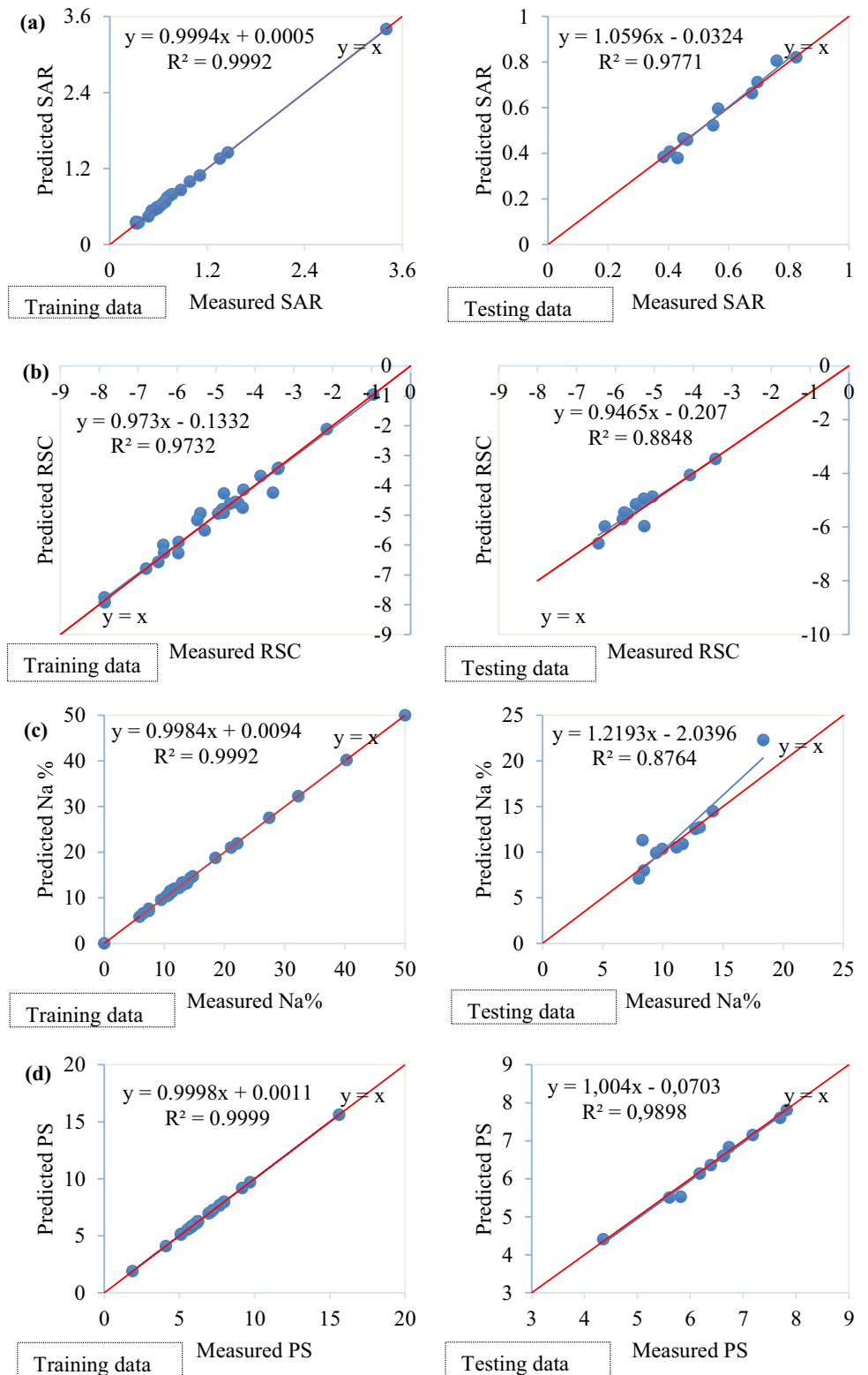
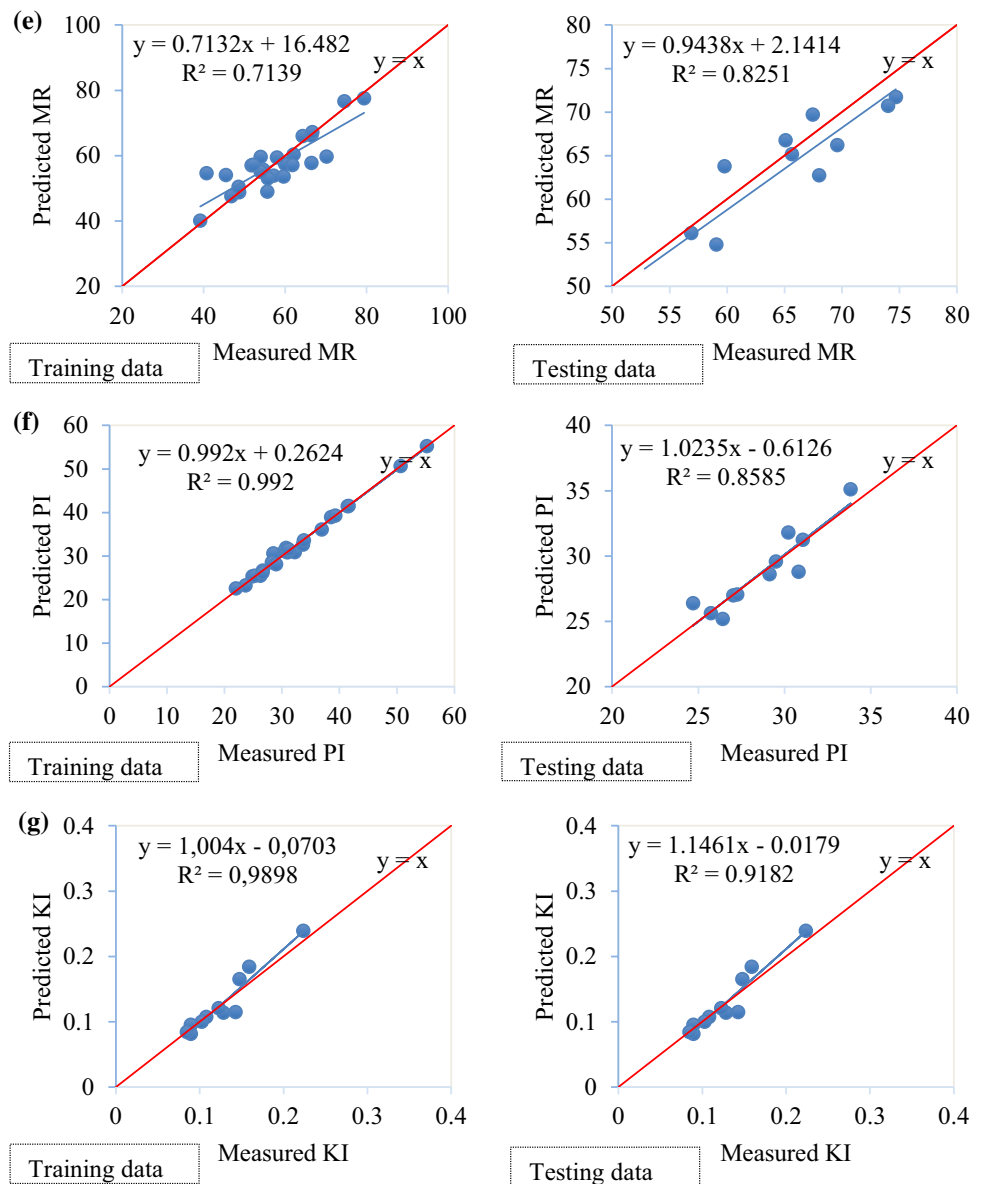


Fig. 9 (continued)



2011). The results showed that the ANN and ANFIS models provide high accuracy for estimating groundwater quality parameters, but the ANN model provides more accuracy than the ANFIS model on average. The possible reason for the better performance of ANN may be because of its ability to learn and recognize linear, nonlinear and complex relations between input and output variables (Mallik et al. 2022). It has been confirmed by El Bilali and Taleb (2020), Yıldız and Karakuş (2020), Ahmed et al. (2019) that artificial intelligence techniques predict water quality with high accuracy.

Scatter plots of measured and ANN model-estimated values of water quality indices in training and testing datasets are presented in Fig. 9. The $y = x$ line indicates the ideal results. The nearer the data points to that line, the

better the model outcomes (Ji et al. 2017). Almost all data points have a close position to $y = x$ line of all scatter plots. The present findings revealed almost a perfect compliance between measured and estimated values.

The result of the present study is in agreement with the findings by M'nassri et al. (2022), Maroufpoor et al. (2020). The methodology used in this study improves the prediction of irrigation water quality parameters. In addition, the use of principal components analysis in the selection of input parameters increases the prediction reliability of the models used (ANN and ANFIS) (Tizro et al. 2021). However, this approach should be tested with water quality data from different regions and at different periods for the same study area.

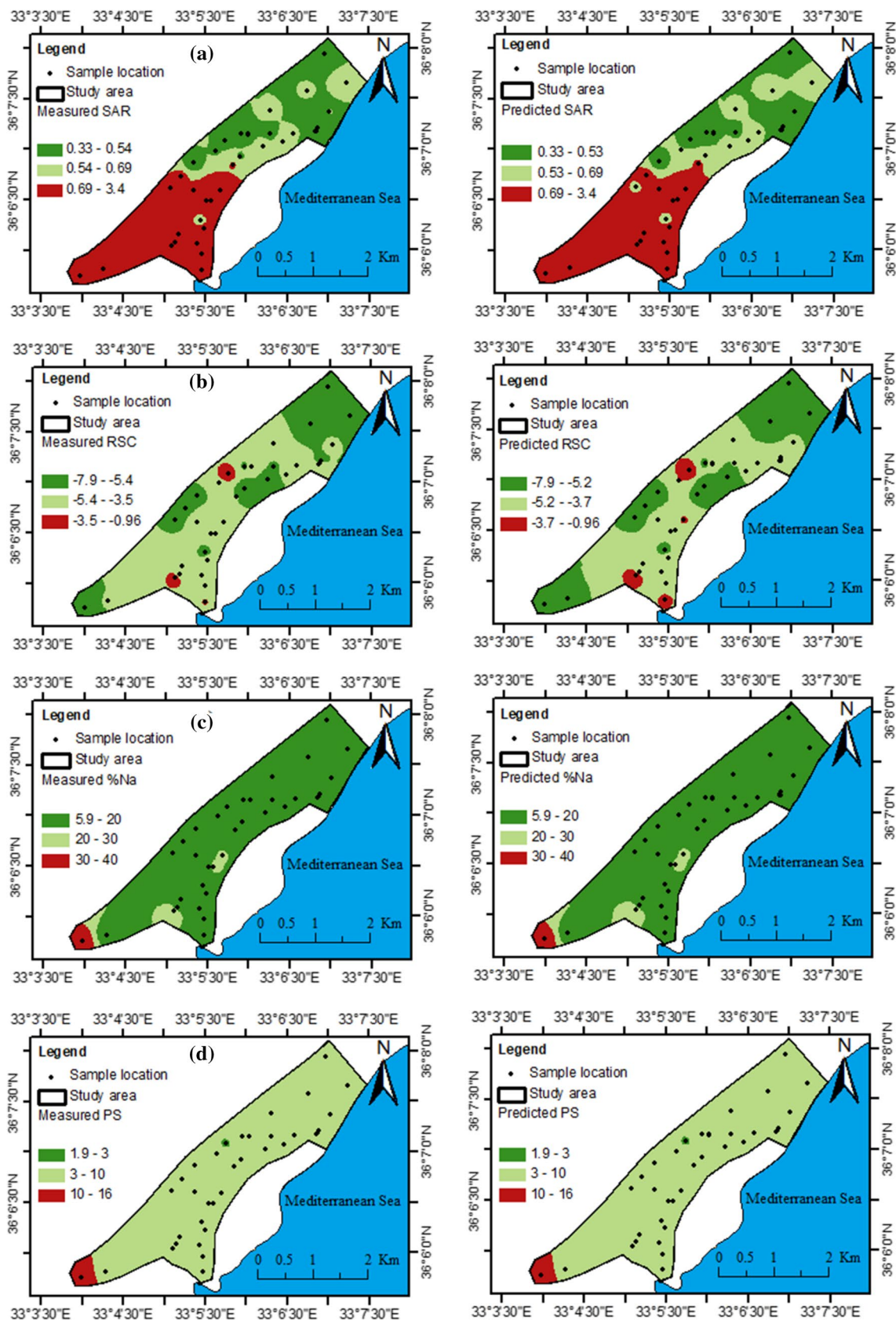


Fig. 10 Spatial distribution of measured and ANN model-estimated optimum values of groundwater quality parameters: **a** SAR, **b** RSC, **c** Na%, **d** PS, **e** MR, **f** PI and **g** KI

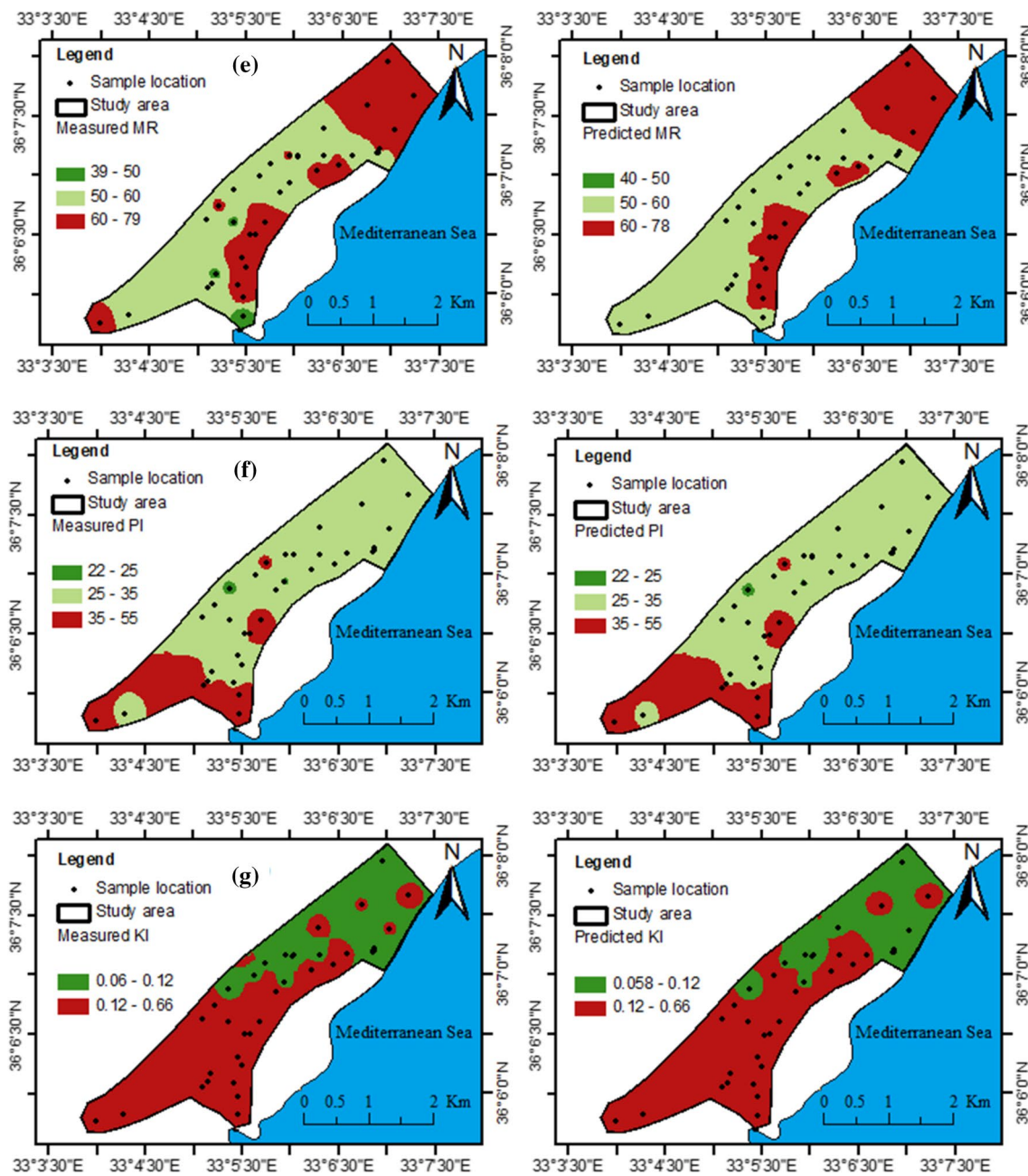


Fig. 10 (continued)

Spatial distribution of groundwater quality indices

SAR is an important criterion for assessing the suitability of water for irrigation since it is the fingerprint of alkaline/sodium threat on crops (Richards 1954; Todd 1959). SAR values of greater than 18 may adversely affect soil structure and plant development (Richards 1954). The spatial distribution of measured and ANN model-estimated SAR values is presented in Fig. 10a. The measured and estimated SAR values had similar distributions, and greater

values were seen in the southeastern section of the study area than the other section. SAR values of the study area varied between 0.33 and 3.4, and since none of the samples had a SAR value of greater than 10, the groundwaters of the study area were found to be suitable for irrigation.

RSC was calculated to determine the harmful effects of carbonate and bicarbonate. High RSC values have highly negative impacts on plant growth and development (Kumar et al. 2009). Tiwari and Manzoor (1988) indicated that negative RSC values emerged when the calcium and magnesium

ions did not precipitate fully. Additionally, negative (–) RSC values indicate the absence of sodium hazard (Eaton 1950). The spatial distribution of measured and ANN model-estimated optimum RSC values is presented in Fig. 10b. The measured and estimated RSC values exhibited quite similar distributions, and greater values were observed locally in the middle of the study area. Since RSC values of the study area were lower than 1.25 meq/L, the present groundwaters were found to be suitable for irrigation.

The sodium of irrigation waters is generally expressed in the percentage of sodium (Na%). It is among the most common parameters used to assess the suitability of groundwaters for irrigation. Sodium cation of water reduces soil permeability; thus, it is not desired in irrigation waters. According to Wilcox (1955), Na% values of up to 60% indicate the suitability of water for irrigation. The present ANN model was the most suitable model for Na%. The spatial distribution of measured and estimated Na% values is presented in Fig. 10c. The measured and estimated Na% maps were almost identical (Fig. 10c). The Na% values of all samples were below 60%. Since the majority of the samples taken from the study area had Na% values less than 20%, they were classified as “perfect” for irrigation. The Na% values were between 20 and 40% within a small section the southwest of the study area, and these waters were classified as “good” for irrigation.

Potential salinity (PS) may increase the osmotic potential of soil solution when the soil moisture is below 50%. It estimates the hazard of high salt concentrations induced by Cl^- and SO_4^{2-} (Güngör and Arslan 2016). The spatial distribution of measured and estimated PS values is presented in Fig. 10d. The measured and estimated PS maps were almost identical. In the majority of the study area, PS values varied between 3 and 10 meq/L; thus, the salinity level was considered as moderate. PS values increased in a small portion of the west of the study area, and potential hazards were expected in future when these waters are used in irrigation.

Magnesium ratio (MR) is also used to assess the suitability of groundwaters for irrigation. Raghunath (1987) indicated that the high magnesium ratio of irrigation water increased soil alkaline levels and negatively influenced crop yield. Thapa et al. (2017) suggested that excessive Mg^{2+} concentrations in irrigation waters reduced soil available K^+ content and had toxic effects on plants. Therefore, MR was used in the present study to assess groundwater quality. Increasing MR values reduce the suitability of waters for irrigation. MR of irrigation waters should be less than 50% (Szabolcs and Darab 1964). The spatial distribution of measured and estimated MR values is presented in Fig. 10e. Despite minor differences, maps were similar primarily with each other. The calculated MR values of the study area varied between 39.2 and 79.4%. The majority of the study area had an MR value of less than 50%. However, the rest had

MR values of greater than 50%; thus, these waters were not considered suitable for irrigation (Fig. 10e).

The permeability index (PI) was developed by Doneen (1964) to assess the suitability of water for irrigation. Soil permeability is influenced by the long-term use of irrigation waters rich in Na^+ , Mg^{2+} , Ca^{2+} , and HCO_3^- . The waters with a PI value of greater than 75 are classified as “Class-1” and considered as the most suitable waters for irrigation; the waters with PI values of between 25 and 75 are classified as “Class-2” and considered as moderately suitable for irrigation; and the waters with a PI value of less than 25 are classified as unsuitable for irrigation (Doneen 1964). The spatial distribution of measured and estimated PI values is presented in Fig. 10f. Spatial distribution maps of measured and estimated PI values were almost identical. In a small portion of the study area, PI values were less than 25 and the groundwaters of these sections were not found to be suitable for irrigation. The rest of the study area had PI values of between 25 and 75 and thus classified as suitable for irrigation.

Kelly (1940) introduced an important parameter, “Kelly Index—KI” to assess the irrigation water quality based on Na^+ level measured against Ca^{2+} and Mg^{2+} . Water Na^+ , Ca^{2+} and Mg^{2+} concentrations represent the alkaline hazard (Dhembare 2012). The waters with a KI value of greater than 1 are unsuitable for irrigation because of excessive Na^+ concentrations and the risk of dispersing soil. On the other hand, KI values lower than 1 are accepted as suitable for irrigation. The spatial distribution of measured and estimated KI values is presented in Fig. 10g. Spatial distribution maps of measured and estimated KI values were similar to each other. KI values of the present groundwaters were lower than 1 and thus considered to be suitable for irrigation. KI values were greater in the southern sections of the study area as compared to the other sections.

Conclusions

In this study, the suitability of groundwater for irrigation purposes was evaluated by using seven different water quality indices. In addition, ANN and ANFIS models were developed and compared to estimate SAR, RSC, %Na, PS, MR, PI and KI parameters by using some physical and chemical properties of groundwater as input. Finally, spatial distribution maps were prepared using GIS for irrigation purpose assessments of the groundwater. Four different statistical parameters (R^2 , RMSE, MAE and NS) were used to compare the performance of ANN and ANFIS models. The input variables to be used in the models were determined according to the results of the correlation analysis and principal component analysis. In the estimation of all irrigation water quality parameters, the ANN model has performed

much higher compared to the ANFIS model. It was observed based on R^2 , RMSE, MAE and NS values that models had better estimation capacity for SAR, RSC, PS, and KI than for Na%, MR, and PI. According to the spatial distribution maps, it was determined that MR was above the allowable limit in the majority of the study area, while PS was high in a small area in the west of the study area. Groundwater in these regions needs to be monitored and managed for irrigation purposes. As a result, the proposed ANN method can be used as an effective tool for groundwater quality assessment in the current study area. The fact that water quality parameters were collected from groundwater wells in the study area and analyzed in the laboratory is a concern and limitation of this study. To improve or validate the findings of the current study, it is recommended to increase the number of sampling points, sampling at different periods, and estimate quality indices with other machine learning models. It should be kept in mind that the present findings were valid for the present study area and should be tested for the other regions.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s13201-022-01810-4>.

Acknowledgements The author would like to thank Dr. Mehmet TAŞAN for his help in the water samplings and laboratory analysis.

Author's contribution The paper is the result of one author's work.

Funding The author did not receive support from any organization for the submitted work.

Declarations

Conflict of interest The author declares no conflict of interest.

Ethical approval Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abdulshahed AM, Longstaff AP, Fletcher S (2015) The application of ANFIS prediction models for thermal error compensation on CNC machine tools. *Appl Soft Comput* 27:158–168. <https://doi.org/10.1016/j.asoc.2014.11.012>
- Abyaneh HZ, Nia AM, Varkeshi MB, Marofi S, Kisi O (2011) Performance evaluation of ANN and ANFIS models for estimating garlic crop evapotranspiration. *J Irrig Drain Eng* 137(5):280–286. [https://doi.org/10.1061/\(ASCE\)IR.1943-4774.0000298](https://doi.org/10.1061/(ASCE)IR.1943-4774.0000298)
- Adamowski J, Chan HF, Prasher SO, Sharda VN (2012) Comparison of multivariate adaptive regression splines with coupled wavelet transform artificial neural networks for runoff forecasting in Himalayan micro-watersheds with limited data. *J Hydroinformatics* 14(3):731–744. <https://doi.org/10.2166/hydro.2011.044>
- Adimalla N, Taloor AK (2020) Hydrogeochemical investigation of groundwater quality in the hard rock terrain of South India using Geographic Information System (GIS) and groundwater quality index (GWQI) techniques. *Groundw Sustain Dev* 10:100288. <https://doi.org/10.1016/j.gsd.2019.100288>
- Adimalla N, Wu J (2019) Groundwater quality and associated health risks in a semi-arid region of south India: Implication to sustainable groundwater management. *Hum Ecol Risk Assess* 25(1–2):191–216. <https://doi.org/10.1080/10807039.2018.1546550>
- Ahmed U, Mumtaz R, Anwar H, Shah AA, Irfan R, García-Nieto J (2019) Efficient water quality prediction using supervised machine learning. *Water* 11(11):2210. <https://doi.org/10.3390/w11112210>
- Al-Waeli LK, Sahib JH, Abbas HA (2022) ANN-based model to predict groundwater salinity: a case study of West Najaf-Kerbala region. *Open Eng* 12(1):120–128. <https://doi.org/10.1515/eng-2022-0025>
- American Public Health Association (APHA) (2005) Standard methods for the examination of water and waste water, 21st edn. American Public Health Association, Washington
- Anonymous (2020) Mersin climatic data Turkish State, Meteorological Service <https://www.mgm.gov.tr/veridegerlendirme/il-ve-ilceler-istatistik.aspx?k=Am=MERSIN>. Accessed 22 Oct 2020
- Arslan H (2017) Determination of temporal and spatial variability of groundwater irrigation quality using geostatistical techniques on the coastal aquifer of Çarşamba Plain, Turkey, from 1990 to 2012. *Environ Earth Sci* 76(1):38. <https://doi.org/10.1007/s12665-016-6375-x>
- Banadkooki FB, Ehteram M, Panahi F, Sammen SS, Othman FB, Ahmed ES (2020) Estimation of total dissolved solids (TDS) using new hybrid machine learning models. *J Hydrol* 587:124989. <https://doi.org/10.1016/j.jhydrol.2020.124989>
- Barik R, Pattanayak SK (2019) Assessment of groundwater quality for irrigation of green spaces in the Rourkela city of Odisha, India. *Groundw Sustain Dev* 8:428–438. <https://doi.org/10.1016/j.gsd.2019.1005>
- Bhunia GS, Keshavarzi A, Shit PK, Omran ESE, Bagherzadeh A (2018) Evaluation of groundwater quality and its suitability for drinking and irrigation using GIS and geostatistics techniques in semi-arid region of Neyshabur, Iran. *Appl Water Sci* 8(6):1–16. <https://doi.org/10.1007/s13201-018-0795-6>
- Chai T, Draxler RR (2014) Root mean square error (RMSE) or mean absolute error (MAE)? Arguments against avoiding RMSE in the literature. *Geosci Model Dev* 7:1247–1250. <https://doi.org/10.5194/gmd-7-1247-2014>
- Chen K, Jiao JJ, Huang J, Huang R (2006) Multivariate statistical evaluation of trace elements in groundwater in a coastal area in Shenzhen, China. *Environ Pollut* 147(3):771–780. <https://doi.org/10.1016/j.envpol.2006.09.002>
- Cigizoglu HK, Kişi Ö (2005) Flow prediction by three back propagation techniques using k-fold partitioning of neural network training data. *Hydrol Res* 36(1):49–64. [https://doi.org/10.2166/nh\(2005\)0005](https://doi.org/10.2166/nh(2005)0005)
- Delgado C, Pacheco J, Cabrera A, Batllori E, Orellana R, Bautista F (2010) Quality of groundwater for irrigation in tropical karst environment: the case of Yucatan, Mexico. *Agric Water Manag* 97:1423–1433. <https://doi.org/10.1016/j.agwat.2010.04.006>
- Dhembare AJ (2012) Assessment of water quality indices for irrigation of Dynaneshwar dam water, Ahmednagar, Maharashtra, India. *Arch Appl Sci Res* 4(1):348–352

- Dinkar KD (2017) Modelling of Reference Evapotranspiration for Western Maharashtra (Doctoral dissertation, Maharana Pratap University of Agriculture and Technology, Udaipur)
- Doneen LD (1964) Water quality for agriculture department of irrigation. University of California, Davis, p 48
- Eaton FM (1950) Significance of carbonates in irrigation waters. *Soil Sci* 69(2):123–134
- El Bilali A, Taleb A (2020) Prediction of irrigation water quality parameters using machine learning models in a semi-arid environment. *J Saudi Soc Agric Sci* 19(7):439–451. [https://doi.org/10.1016/j.jssas\(2020\)08001](https://doi.org/10.1016/j.jssas(2020)08001)
- El Bilali A, Taleb A, Brouziyne Y (2021) Groundwater quality forecasting using machine learning algorithms for irrigation purposes. *Agric Water Manag* 245:106625. <https://doi.org/10.1016/j.agwat.2020.106625>
- Elzain HE, Chung SY, Park KH, Senapathi V, Sekar S, Sabarathinam C, Hassan M (2021) ANFIS-MOA models for the assessment of groundwater contamination vulnerability in a nitrate contaminated area. *J Environ Manage* 286:112162. <https://doi.org/10.1016/j.jenvman.2021.112162>
- Ghazaryan K, Movsesyan H, Gevorgyan A, Minkina T, Sushkova S, Rajput V, Mandzhieva S (2020) Comparative hydrochemical assessment of groundwater quality from different aquifers for irrigation purposes using IWQI: a case-study from Masis province in Armenia. *Groundw Sustain Dev* 11:100459. <https://doi.org/10.1016/j.gsd.2020.100459>
- Gholami V, Khaleghi MR, Pirasteh S, Booij MJ (2022) Comparison of self-organizing map, artificial neural network, and co-active neuro-fuzzy inference system methods in simulating groundwater quality: geospatial artificial intelligence. *Water Resour Manag* 36(2):451–469. <https://doi.org/10.1007/s11269-021-02969-2>
- Güngör A, Arslan H (2016) Assessment of water quality in drainage canals of Çarşamba Plain, Turkey, through water quality indexes and graphical methods. *Glob Nest J* 18(1):67–78
- Haykin S (1999) *Neural networks: a comprehensive foundation*, 2nd edn. Prentice-Hall, Englewood Cliffs
- Hornik K, Stinchcombe M, White H (1989) Multilayer feedforward networks are universal approximators. *Neural Netw* 2:359–366. [https://doi.org/10.1016/0893-6080\(89\)90020-8](https://doi.org/10.1016/0893-6080(89)90020-8)
- Hossain M, Patra PK, Begum SN, Rahaman CH (2020) Spatial and sensitivity analysis of integrated groundwater quality index towards irrigational suitability investigation. *Appl Geochem* 123:104782. <https://doi.org/10.1016/j.apgeochem.2020.104782>
- Jahani A, Feghhi J, Makhdom MF, Omid M (2016) Optimized forest degradation model (OFDM): an environmental decision support system for environmental impact assessment using an artificial neural network. *J Environ Plan Manag* 59(2):222–244. <https://doi.org/10.1080/0964056820151005732>
- Jahin HS, Abuzaid AS, Abdellatif AD (2020) Using multivariate analysis to develop irrigation water quality index for surface water in Kafr El-Sheikh Governorate. *Egypt Environ Technol Innov* 17:100532. [https://doi.org/10.1016/j.jeti\(2019\)100532](https://doi.org/10.1016/j.jeti(2019)100532)
- Jang JS (1993) ANFIS: adaptive-network-based fuzzy inference system. *IEEE Trans Syst Man Cybern* 23(3):665–685. <https://doi.org/10.1109/21.256541>
- Kelly WP (1940) Permissible composition and concentration of irrigated waters. *Proc ASCF* 607:607–613
- Kumar SK, Rammohan V, Sahayam JD, Jeevanandam M (2009) Assessment of groundwater quality and hydrogeochemistry of Manimuktha River basin, Tamil Nadu, India. *Environ Monit Assess* 159(1–4):341. <https://doi.org/10.1007/s10661-008-0633-7>
- Lanjwani MF, Khuhawar MY, Jahangir Khuhawar TM (2020) Assessment of groundwater quality for drinking and irrigation uses in taluka Ratodero, district Larkana, Sindh, Pakistan. *Int J Environ Anal Chem*. <https://doi.org/10.1080/0306731920201780222>
- Li J, Heap AD (2011) A review of comparative studies of spatial interpolation methods in environmental sciences: performance and impact factors. *Ecol Inform* 6(3–4):228–241. <https://doi.org/10.1016/j.ecoinf.2010.12.003>
- Liu CW, Lin KH, Kuo YM (2003) Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan. *Sci Total Environ* 313(1–3):77–89. [https://doi.org/10.1016/S0048-9697\(02\)00683-6](https://doi.org/10.1016/S0048-9697(02)00683-6)
- Longley PA, Goodchild MF, Maguire DJ, Rhind DW (2010) *Geographic information systems and science*, 3rd edn. Wiley
- Mallick J, Singh CK, AlMesfer MK, Kumar A, Khan RA, Islam S, Rahman A (2018) Hydro-geochemical assessment of groundwater quality in Aseer Region, Saudi Arabia. *Water* 10(12):1847. <https://doi.org/10.3390/w10121847>
- Maroufpoor S, Jalali M, Nikmehr S, Shiri N, Shiri J, Maroufpoor E (2020) Modeling groundwater quality by using hybrid intelligent and geostatistical methods. *Environ Sci Pollut Res* 27:28183–28197. <https://doi.org/10.1007/s11356-020-09188-z>
- Mishra AK, Desai VR (2006) Drought forecasting using feedforward recursive neural network. *Ecol Model* 198(1–2):127–138. <https://doi.org/10.1016/j.ecolmodel.2006.04.017>
- M'nassri S, El Amri A, Nasri N, Majdoub R, (2022) Estimation of irrigation water quality index in a semi-arid environment using data-driven approach. *Water Supply* 22(5):5161–5175. <https://doi.org/10.2166/ws2022157>
- Mokoena P, Kanyerere T, van Bever DJ (2020) Hydrogeochemical characteristics and evaluation of groundwater quality for domestic and irrigation purposes: a case study of the Heuningnes Catchment, Western Cape Province, South Africa. *SN Appl Sci* 2(9):1–12. <https://doi.org/10.1007/s42452-020-03339-0>
- Mokhtar A, El-Ssawy W, He H, Al-Ansari N, Sammen SS, Gyasi-Agyei Y, Abuarab M (2022) Using machine learning models to predict hydroponically grown lettuce yield. *Front Plant Sci* 13:706042
- Mondal NC, Singh VP, Singh VS, Saxena VK (2010) Determining the interaction between groundwater and saline water through groundwater major ions chemistry. *J Hydrol* 388(1–2):100–111. <https://doi.org/10.1016/j.jhydrol.2010.04.032>
- Muttill N, Chau KW (2007) Machine-learning paradigms for selecting ecologically significant input variables. *Eng Appl Artif Intell* 20(6):735–744. <https://doi.org/10.1016/j.engappai.2006.11.016>
- Nash JE, Sutcliffe JV (1970) River flow forecasting through conceptual models part I-A discussion of principles. *J Hydrol* 10(3):282–290
- Rauf AU, Ghuman AR, Ahmad S, Hashmi HN (2018) Performance assessment of artificial neural networks and support vector regression models for stream flow predictions. *Environ Monit Assess* 190:704. <https://doi.org/10.1007/s10661-018-7012-9>
- Ravichandra R, Chandana OS (2006) Study on evaluation on ground water pollution in Bakkannapalem, Visakhapatnam. *Nat Environ Pollut* 5(2):203–207
- Richards LA, 1954 Diagnosis and improvement of saline and alkali soils. In: *Agricultural Handbook*, vol 60. USDA and IBH Pub Coy Ltd, New Delhi, pp 98–99
- Sahour H, Gholami V, Vazifedan M (2020) A comparative analysis of statistical and machine learning techniques for mapping the spatial distribution of groundwater salinity in a coastal aquifer. *J Hydrol* 591:125321. <https://doi.org/10.1016/j.jhydrol.2020.125321>
- Sarkar M, Pal SC, Islam ARM (2022) Groundwater quality assessment for safe drinking water and irrigation purposes in Malda district, Eastern India. *Environ Earth Sci* 81(2):1–20. <https://doi.org/10.1007/s12665-022-10188-0>
- Shu C, Ouarda TBMJ (2008) Regional flood frequency analysis at ungauged sites using the adaptive neuro-fuzzy inference system.

- J Hydrol 349(1–2):31–43. <https://doi.org/10.1016/j.jhydrol.2007.07.005>
- Sreekanth PD, Sreedevi PD, Ahmed S, Geethanjali N (2011) Comparison of FFNN and ANFIS models for estimating groundwater level. *Environ Earth Sci* 62(6):1301–1310. <https://doi.org/10.1007/s12665-010-0617-0>
- Szabolcs I, Darab C (1964) The influence of irrigation water of high sodium carbonate content on soils. In: Szabolcs I (ed) Proc 8th international congress soil science sodics soils, res inst soil sci agric chem Hungarian acad sci, ISSS Trans II, pp 802–812
- Thapa R, Gupta S, Reddy DV, Kaur H (2017) An evaluation of irrigation water suitability in the Dwarka river basin through the use of GIS-based modelling. *Environ Earth Sci* 76(14):471. <https://doi.org/10.1007/s12665-017-6804-5>
- Tiwari TN, Manzoor A (1988) Water quality index for Indian rivers. In: Trivedy RK (ed) Ecology and pollution of Indian rivers. Aashish Publishing House, New Delhi, pp 271–286
- Tiyasha, Tung TM, Yaseen ZM (2020) A survey on river water quality modelling using artificial intelligence models: 2000–2020. *J Hydrol* 585:124670. <https://doi.org/10.1016/j.jhydrol.2020.124670>
- Tizro AT, Fryar AE, Vanaei A, Kazakis N, Voudouris K, Mohammadi P (2021) Estimation of total dissolved solids in Zayandehrood River using intelligent models and PCA. *Sustain Water Resour Manag* 7(2):1–13
- Todd DK (1959) Groundwater hydrology. Wiley, p 535
- Trabelsi F, Bel Hadj Ali S (2022) Exploring machine learning models in predicting irrigation groundwater quality indices for effective decision making in Medjerda River Basin, Tunisia. *Sustainability* 14(4):2341. <https://doi.org/10.3390/su14042341>
- Wanas N, Auda G, Kamel MS, Karray F (1998) On the optimal number of hidden nodes in a neural network. In Conference proceedings IEEE Canadian conference on electrical and computer engineering, vol 2, pp 918–921
- Wilcox LV (1955) Classification and use of irrigation waters. US Dept of Agric, Circular No 696, Washington, p 19
- Yıldız S, Karakuş CB (2020) Estimation of irrigation water quality index with development of an optimum model: a case study. *Environ Dev Sustain* 22(5):4771–4786. <https://doi.org/10.1007/s10668-019-00405-5>
- Yu H, Wen X, Wu M, Sheng D, Wu J, Zhao Y (2022) Data-based groundwater quality estimation and uncertainty analysis for irrigation agriculture. *Agric Water Manag* 262:107423. <https://doi.org/10.1016/j.agwat.2021.107423>

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