

Regular Article

Luay Kadhim Al-Waeli*, Jumana Hadi Sahib and Hiba Ali Abbas

ANN-based model to predict groundwater salinity: A case study of West Najaf–Kerbala region

<https://doi.org/10.1515/eng-2022-0025>

received August 22, 2021; accepted December 10, 2021

Abstract: Estimating groundwater salinity is important for the use of groundwater resources for irrigation purposes and provides a suitable guide for the management of groundwater. In this study, the artificial neural networks (ANNs) were adopted to estimate the salinity of groundwater identified by total dissolved solids (TDS), sodium adsorption ratio (SAR), and sodium (Na^+) percent, using electrical conductivity, magnesium (Mg^{2+}), calcium (Ca^{2+}), potassium (K^+), and potential of hydrogen (pH) as input elements. Samples of groundwater were brought from 51 wells situated in the plateau of Najaf–Kerbala provinces. The network structure was designed as 6-4-3 and adopted the default scaled conjugate gradient algorithm for training using SPSS V24 software. It was observed that the proposed model with four neurons was exact in estimating the irrigation salinity. It has shown a suitable agreement between experimental and ANN values of irrigation salinity indices for training and testing datasets based on statistical indicators of the relative mean error and determination coefficient R^2 between ANN outputs and experimental data. TDS, SAR, and Na percent predicted output tracked the measured data with an R^2 of 0.96, 0.97, and 0.96 with relative error of 0.038, 0.014, and 0.021, respectively, for testing, and R^2 of 0.95, 0.96, and 0.96 with relative error of 0.053, 0.065, and 0.133, respectively, for training. This is an indication that the designed network was satisfactory. The model could be utilized for

new data to predict the groundwater salinity for irrigation purposes at the Najaf–Kerbala plateau in Iraq.

Keywords: groundwater quality, Najaf–Kerbala plateau, artificial neural network

1 Introduction

Recently, drought has been a vital problem globally due to a lack of amounts precipitation. It affected the provision of freshwater sources. To compensate for the shortfall, it seemed to rely on drilling wells. It also became necessary to monitor the quality of the groundwater extracted from these wells. The monitoring process is carried out by taking samples of healthy water and testing them in a laboratory. This is a time-consuming process. Recently, an artificial neural network (ANN) method was used to estimate the quality of the groundwater to save time and effort. ANN is a mathematical technique that relies on model training to predict the quality of water extracted from wells [1].

Many researchers have worked on groundwater quality, where Ghalib et al. [2] predicted the Bahar Al-Najaf basin's water quality by adopting traditional methods. The calculation was based on the main and secondary elements such as calcium (Ca), sodium (Na), potassium (K), magnesium (Mg), chloride (Cl), sulfate (SO_4), nitrate (NO_3), bicarbonate (HCO_3), and bromine (Br), and field measurements include temperature (T), the potential of hydrogen (pH), electrical conductivity (EC), total dissolved solids (TDSs). They found that the water quality was not usable. Banerjee et al. [3] built a model of ANNs to predict salinity at various pumping rates. The quick propagation algorithm was adopted in the feed-forward ANN model. They found that the model is simpler and more accurate than the numerical methods. Khudair et al. [4] developed a model for groundwater quality assessment based on the water quality index. They used four elements, pH, Cl, SO_4 , and TDS, in the statistical program (SPSS) to train ANNs. They produced a neural

* **Corresponding author: Luay Kadhim Al-Waeli**, Civil Engineering Department, University of Kufa, Faculty of Engineering, Najaf Governorate, Kufa, Iraq, e-mail: luayk.alwaeli@uokufa.edu.iq

Jumana Hadi Sahib: Civil Engineering Department, University of Kufa, Faculty of Engineering, Najaf Governorate, Kufa, Iraq, e-mail: Jumanah.alsabarawi@uokufa.edu.iq

Hiba Ali Abbas: Structures and Water Resources Engineering, University of Kufa, Faculty of Engineering, Najaf Governorate, Kufa, Iraq, e-mail: Hiba.abbas@uokufa.edu.iq

network that predicted high efficiency. The pH and Cl were the most influential elements. Heidarzadeh [1] used Na as the response variable in ANNs. He trained the ANN with three inputs such as pH, EC, and total hardness (TH) and found that the ANN of a two-layer of the Logsig-Tansig transfer function with four and three neurons in the first and second layers was the best method compared with others. Kheradpisheh et al. [5] developed an ANN in the MATLAB program and used chemical elements, evaporation, water elevation, and discharge as input. After training, the network extracted Cl, EC, NO_3 , and SO_4 . It gave an excellent response to Cl, EC, and NO_3 , but it was not efficient enough to predict SO_4 because they believe in the various water sources or the influence of other factors. Kılçaslan et al. [6] designed a model to predict groundwater quality using ANNs remotely. Sodium adsorption ratio (SAR) and SO_4 are the outcome for network training on elements pH, EC, TDS, Cl, and TH with two hidden layers. It gave the network efficiency and accuracy. Sunayana et al. [7] used ANNs to predict the quality of water in a landfill. They used three algorithms to train the model, and the best model was based on prediction. They used elements such as pH, TDS, chlorides Cl^- , NO_3^- and SO_4^{2-} with time and locations in the input layer, and the output layer had one element, the total hardness. They used analysis sensitivity to find the elements that most affect the total hardness. They drew a map using ArcGIS, showing the prediction of water quality for the study area. Aryafar et al. [8] used several models of artificial intelligence in addition to genetic programming (GP) to determine the water quality of 12 wells in the Khezri plain, eastern Iran. They used GP, ANN, and adaptive neuro-fuzzy inference system (ANFIS) models. They adopted the outputs such as TH, TDS, and EC and the values of chemical elements as inputs to evaluate the models from the statistical analysis. Where model GP gave the best result than ANN and ANFIS. ANN and ANFIS gave satisfactory results. Maroufpoor et al. [9] depended on hybrid artificial intelligence by merging several algorithms, resulting in the following models: ANN, ANN-PSO, NF-GP, and NF-FCM. They compared hybrid models with geostatistical models as inverse distance weighting, radial basis function, and kriging. Through sensitive analysis for the models, they concluded that the NF-GP model is the best to determine the quality of groundwater in wells by calculating EC, SAR, and Cl. Bedi et al. [10] depended on ANN, support vector machines, and extreme gradient boosting to evaluate three levels of pollution in groundwater in 203 wells spreading across the Midwest in the USA through nonlinear relationships. Land use, hydrological data, and water quality were based on concentrations of nitrates and pesticides.

In this study, the application of computer skills will be used instead of traditional methods to be sensitive

to prediction errors. The model will give an idea of the sensitivity of each of the elements included in the study. The current study aims to use ANN to develop a model to estimate the value of SAR, Na percent, and TDS, which reflects the extent of groundwater salinity in the selected region. Mg, Ca, K, Na, Mg, EC, and pH are used as inputs in the ANN.

2 Description of the study area

The study region is situated in the north-western zone of the Euphrates region, which is the western zone of Mesopotamia, as it is between Najaf and Kerbala governorate. It is located geographically on the longitude $31055'-32045'$ and latitude $43030'-44030'$. It is an almost flat conical plateau with about $2,700 \text{ km}^2$ as in Figure 1. It is bordered to the south by Tar al Al-Najaf, to the north-east by Tar El-Sayyed, to the north via Lake Razzazah, and to the east by quadruple sediments. Its topography elevation is from 13 to 207 m with a mean of 83 m. The elevation decreases from west to east. Shallow valleys cover its surface with flat floors. Its soil is either gravel, gypsum, concrete, or rocky. The study region is stable, and the sedimentary covers range from 7 to 8 km above the basement of rocks, hydrology the Dabdaba.

The region is one of the five layers that make up the plateau, and it is the principal unconfined aquifer in the study region, where it covers an area $1,100 \text{ km}^2$ of the studied area. This aquifer is directly recharged via seasonal rainfalls inside the plateau. The flow direction of the groundwater in the Dabdaba reservoir is toward the Euphrates river (from the southwest to the northeast). The hydraulic grade ranges from 0.0011 to 0.0005, as shown in Figure 2. The General Commission drilled the wells for groundwater, Kerbala Branch. The well water samples were collected by Al-kubaisi et al. [11] in the period April–May, which is considered a wet period, where Cl, NO_3 , SO_4 , HCO_3 , anion, cation, TH, SAR, Na percent, K, Mg, and Ca were measured in the general groundwater authority laboratories. As for temperature, EC, TDS, DO, and pH, it was measured in the site.

3 ANNs modeling

ANNs are one of the mathematical techniques that is simulator of the human brain in its work. Its work depends on internal connections called neurons. The nodes deal with inputs and outputs to know the relationship between them. Relationships are either linear or nonlinear and with complex interconnectedness. The ANN contains three

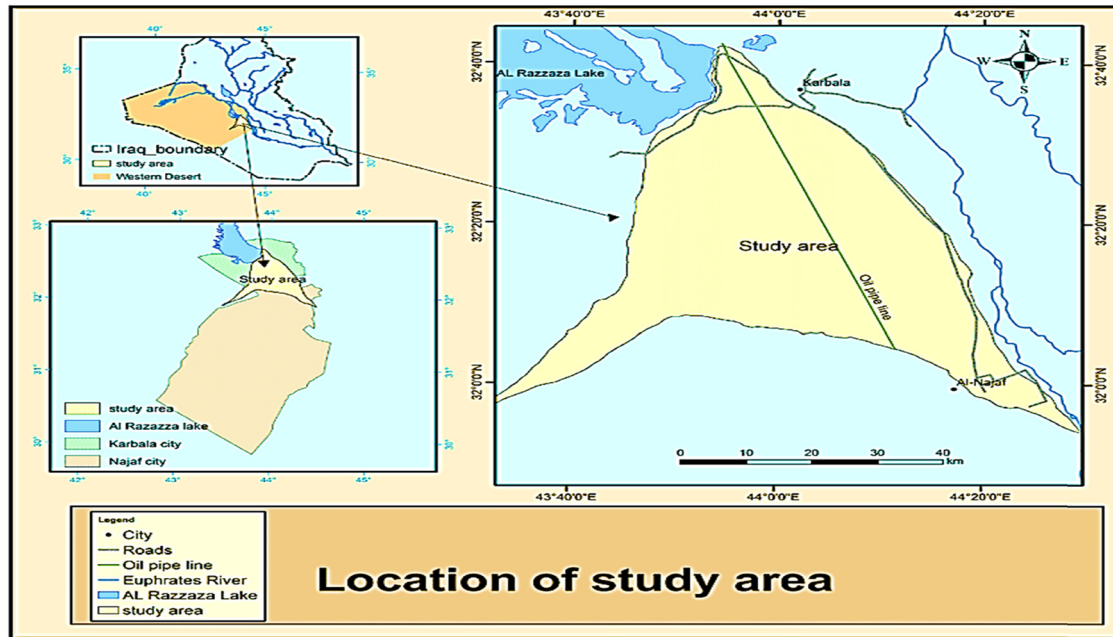


Figure 1: Location of study area (Najaf and Kerbala plateau; Al-kubaisi et al., [11]).

layers: the input, the hidden, and the output. Each layer consists of a group of nerve nodes that connect these layers (which are called neurons) with a complex network and different weights [1] and [6]. The weights give an impression of the strength and influence of each of the inputs on the outputs [12].

In hydrogeology and water resources, this process has been successfully applied. Usually, the nonlinear relationship is used [11]. There are also many algorithms to train the network. It has a quick convergence solution. Unique learning rates are used for each type of contact weight, which is adjusted dynamically by the training process [3]. The weights of neurons in the ANN are connected between the input and the output layer parameters, which determine the pattern of the relationship between them and the bias and algorithm of the middle layer. The hidden layer is usually responsible for the network training process. The hidden layer can consist of more than one layer. The difficulty of the network is characterized by the number of hidden layers and nervous seats. After entering data in the first layer, it is processed in the next layer as output for the first layer. Finally, it is transferred to the output layer after it is accepted. The process is represented by ref. [13]:

$$y_k = f_o \left[\sum_{i=1}^{M_N} W_{ki} \times f_h \left(\sum_{i=1}^{N_N} W_{ji} \times X_i + \omega_{j0} \right) + \omega_{k0} \right], \quad (1)$$

where W_{ji} is a weight of hidden layer linking between the neuron node in the input layer (i th) and the hidden layer

(j th); ω_{j0} is a hidden neuron bias (j th); f_h is an activation algorithm of the hidden neuron; W_{ki} is a weight of output layer linking between the neuron node in the hidden layer (j th) and the output layer (k th); ω_{k0} is an output neuron bias (k th); f_o is the activation algorithm for the output neuron; X_i is an input parameter in the input layer (i th); y_j is an output variable calculated; N_N and M_N are the numbers of node neurons in the hidden and input layers, respectively.

Data were normalized before training and network testing to ensure speedy data processing and reduce error forecasting and convergence during training. As many algorithms and functions were modified, initial data for inputs and outputs were normalized before training and tested by equation (2). Converting the value of the outputs to the actual value after completing the training and normalizing according to equation (3) [12,13].

$$\hat{X} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}, \quad (2)$$

$$X = \hat{X} \times (X_{\max} - X_{\min}) + X_{\min}, \quad (3)$$

where X is the original value; X_{\max} and X_{\min} are maximum and minimum values in the sequence; \hat{X} is the normalized data.

The work of the ANN is evaluated by calculating the coefficient of determination (R^2) and the relative mean error (RME) to determine its accuracy. The cumulative error of the model is accepted through observations and calculations as the lowest value of RME and the highest

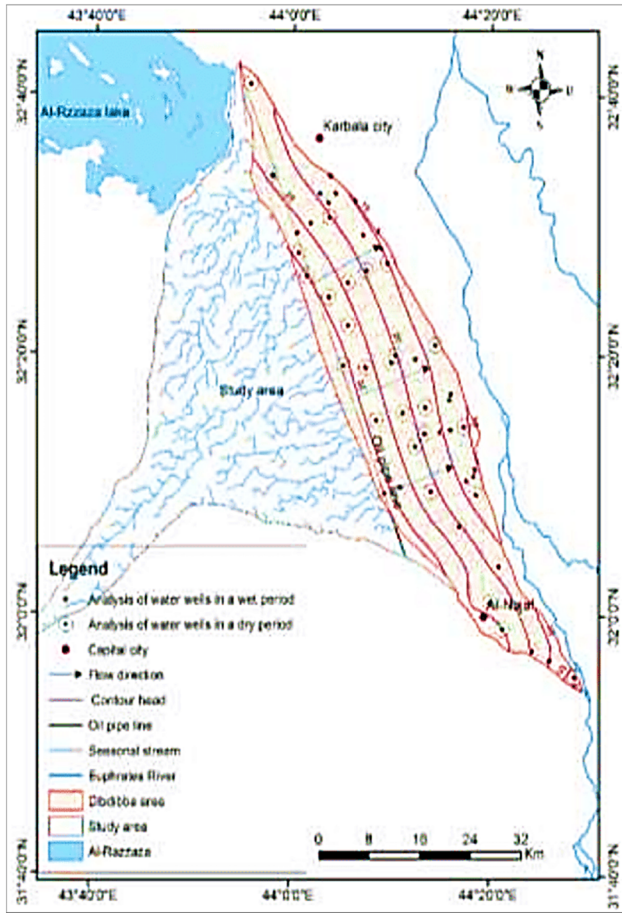


Figure 2: Location of wells in Najaf and Kerbala plateau (Al-kubaisi et al. [11]).

value of R^2 . Simulation work was evaluated in the calculation of the outputs, training, and testing by goodness fit measurement.

$$R^2 = \left[\sum_{i=1}^N (P_i - \bar{P})(O_i - \bar{O}) \right]^2 \left[\sum_{i=1}^N (P_i - \bar{P})^2 (O_i - \bar{O})^2 \right]^{-1}, \quad (4)$$

$$\text{RME} = \left[N^{-1} \sum_{i=1}^N (P_i - O_i) \right], \quad (5)$$

where N is number of observations; P_i is the forecast values; O_i is the observed data; \bar{P} and \bar{O} are the mean values for P_i and O_i , respectively.

In this study, 51 wells in the region Najaf–Kerbala plateau were used for the network model. Around 70% of the wells were approved to train the network, and 30% to test the efficiency of the network based on (214) and (92) readings recorded respectively. The optimal model chosen based on the first step is to determine the type of the transfer function and the number of layers. Input

variables keep different. The data (EC, Ca^{2+} , Mg^{2+} , Na^+ , K, and pH) will be entered as variable inputs to the network and each well. By comparing the highest value of R^2 and the lowest value of RME with the results of the models, the best network of variables was chosen. After selecting the network, they tested the network for the specific layers and model performance through the different transport functions in the second step, and 30% were approved to determine this step.

4 Methodology

4.1 Groundwater data

Samples of groundwater were brought from 51 observation wells situated in the plateau of the Najaf–Kerbala provinces, Iraq. The samples were collected by Al-kubaisi et al. [11] during April–May 2018, which is considered as a wet period. The applicability of ANN was examined to estimate the salinity indices TDS, SAR, and Na percent in 51 observation wells in the Najaf–Kerbala plateau. In this study, Table 1 presents the input elements used in the prediction of salinity indices. The variation of the input element is provided in Figure 3:

4.2 ANN modeling

ANNs are characterized as systems of linked nodes that can simulate parameters from input data. The structure of the ANN is posed from several neurons joined by links and arranged in layers. A layer is entirely connected to the preceding one through weights. Primarily proposed weights are gradually adjusted through the process of

Table 1: Statistical values of the input data

| Elements | Maximum | Minimum | Average | Standard deviation |
|--------------------------------|---------|---------|---------|--------------------|
| Na^+ | 44.58 | 5.916 | 19.72 | 8.56 |
| EC ($\mu\text{S}/\text{cm}$) | 10750 | 3750 | 5136.94 | 1607.42 |
| Ca^{+2} | 30.54 | 6.737 | 13.01 | 4.94 |
| Mg^{+2} | 25.1 | 4.937 | 10.38 | 3.85 |
| pH | 7.61 | 7.1 | 7.18 | 0.12 |
| K^+ | 2.94 | 0.08 | 1.20 | 0.92 |

All values expressed in ppm except EC and pH.

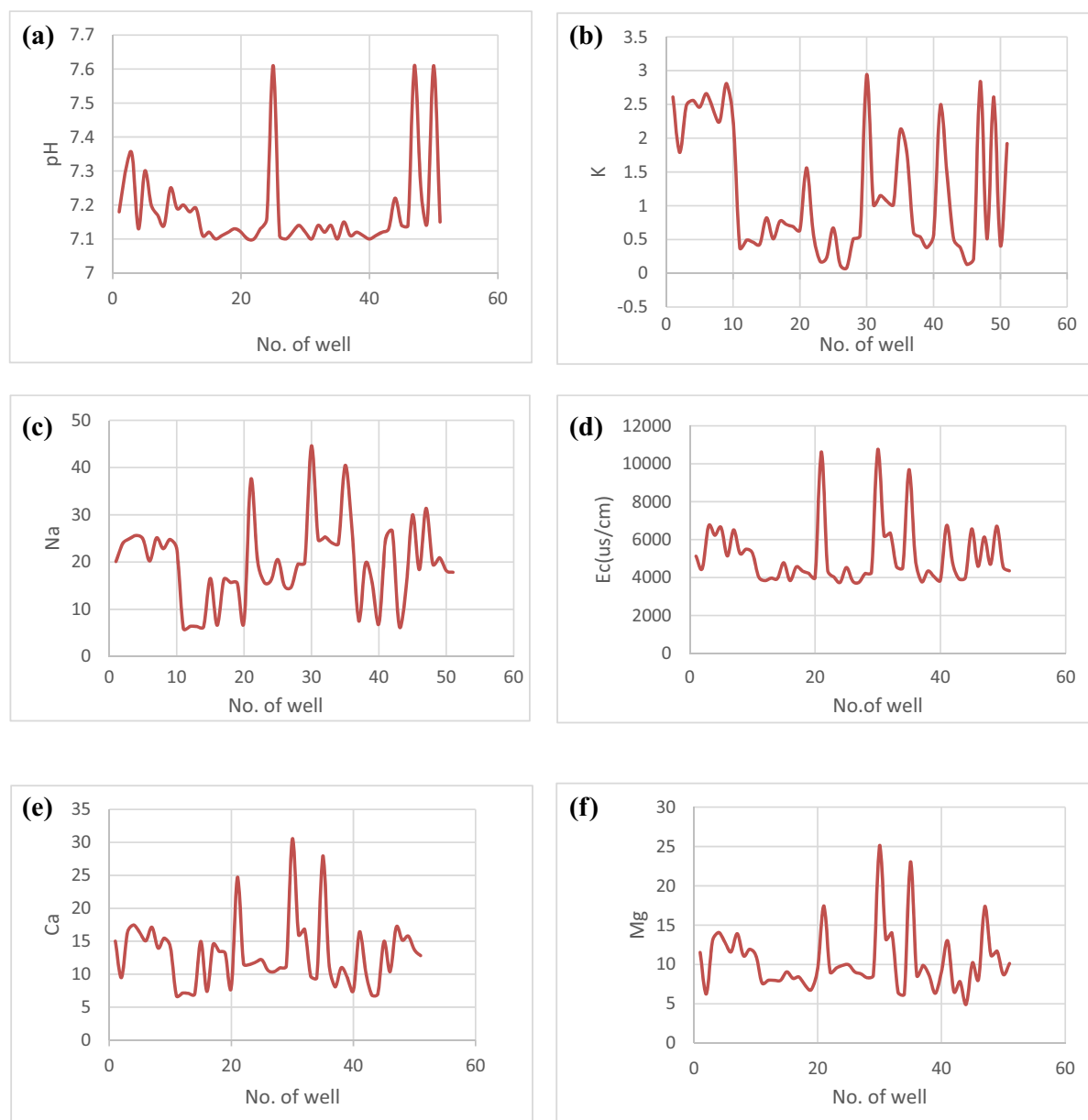


Figure 3: The variation of input variables in ANNs model. (a) The measured data of pH from observation wells. (b) The measured data of K from observation wells. (c) The measured data of Na from observation wells. (d) The measured data of EC from observation wells. (e) The measured data of Ca from observation wells. (f) The measured data of Mg from observation wells.

training via comparing measured data with simulated outputs. The calculation of biases and weights is identified as training. The training process aims to capture the optimal weights by minimizing the error of the output parameters. At once, functions of learning are utilized to update layer's bias and weight. Then, a data set is used randomly to investigate the generalization of a network, which is known as the test stage [14].

In the present study, a back-propagation algorithm training in the training step was adopted to relate the association between input elements EC, Na^+ , Ca^{2+} , Mg^{2+} ,

K^+ , and pH and output salinity expressed in TDS, SAR, and Na percent. The configuration of ANNs was achieved via adopting numerous trials till good results were reached. The ANN plot network properties are shown in Figure 4.

The suitability of the proposed ANN model is assessed by adopting the determination coefficient (R^2) defined based on output prediction errors and the values of R^2 . The condition of acceptance depends on the quantitative error through the predicted and measured data, including maximum R^2 and minimum RME [15]. The network information is listed in Table 2.

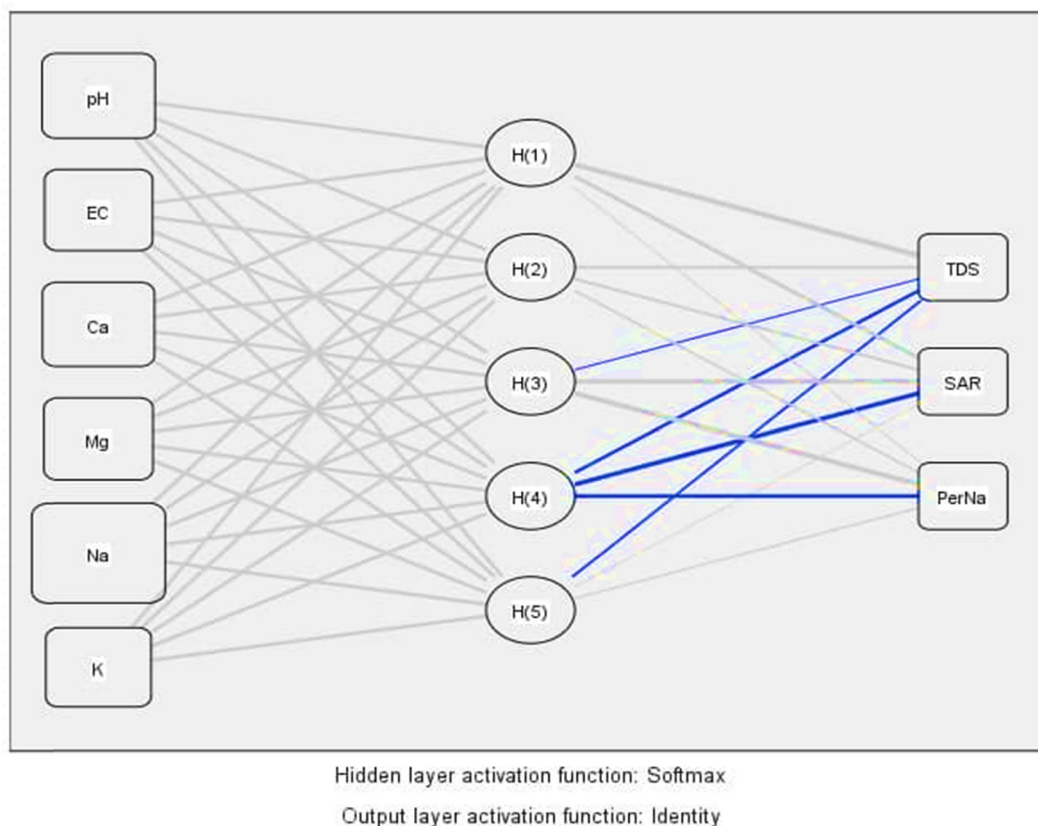


Figure 4: The proposed ANN.

Table 2: The network information for proposed ANN

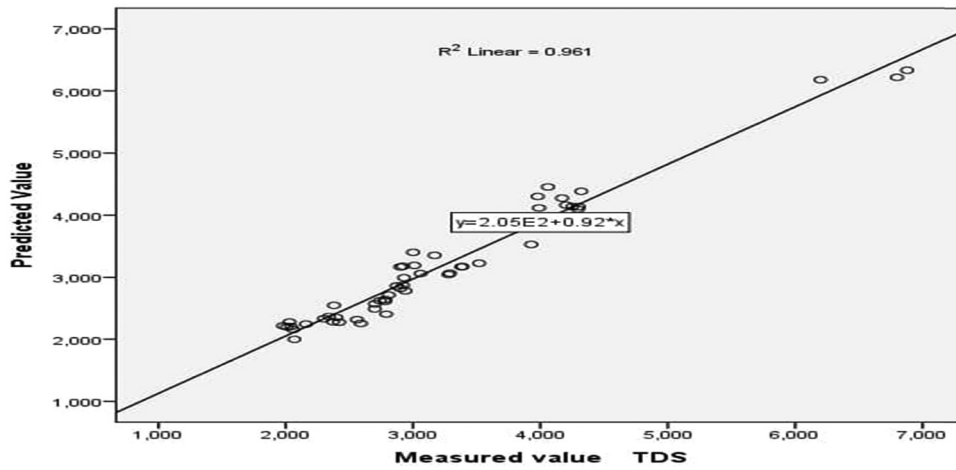
| | | | |
|-----------------|--|-----------------|--------------------|
| Input layer | Input variables | 1 | pH |
| | | 2 | EC |
| | | 3 | Ca |
| | | 4 | Mg |
| | | 5 | Na |
| | | 6 | K |
| | | Number of units | 6 |
| | Rescaling technique for covariates | | Standardized |
| Hidden layer(s) | No. of hidden layers | 1 | |
| | No. of units in hidden layer 1a | 4 | |
| | Algorithm of activation | | Hyperbolic tangent |
| Output layer | Dependent elements | 1 | TDS |
| | | 2 | SAR |
| | | 3 | Na% |
| | No. of units | 3 | |
| | Rescaling technique for scale dependents | | Standardized |
| | Algorithm of activation | | Identity |
| | Method of error | | Sum of squares |

Table 3: The statistical results of the ANNs model

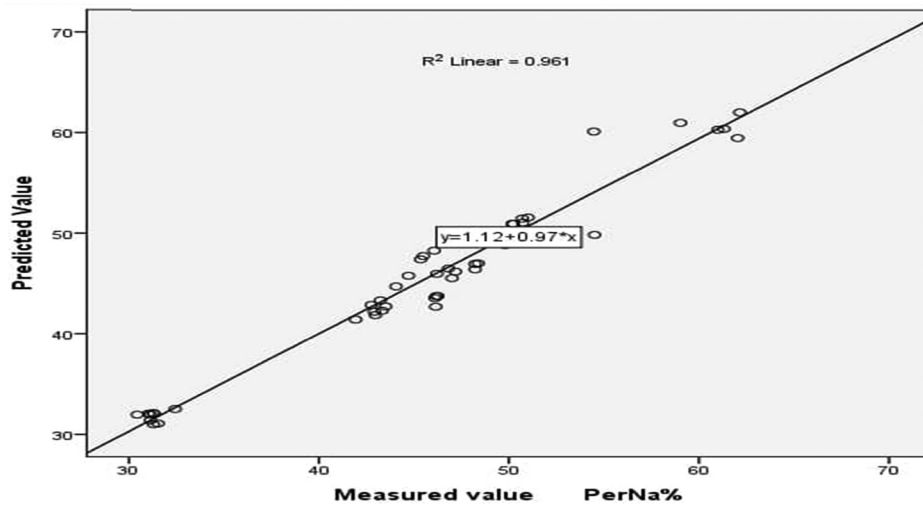
| | | Relative error | R^2 |
|----------|-----|----------------|-------|
| Training | TDS | 0.038 | 0.95 |
| | SAR | 0.014 | 0.96 |
| | Na% | 0.021 | 0.96 |
| Testing | TDS | 0.053 | 0.96 |
| | SAR | 0.065 | 0.97 |
| | Na% | 0.113 | 0.96 |

5 ANNs model results and discussion

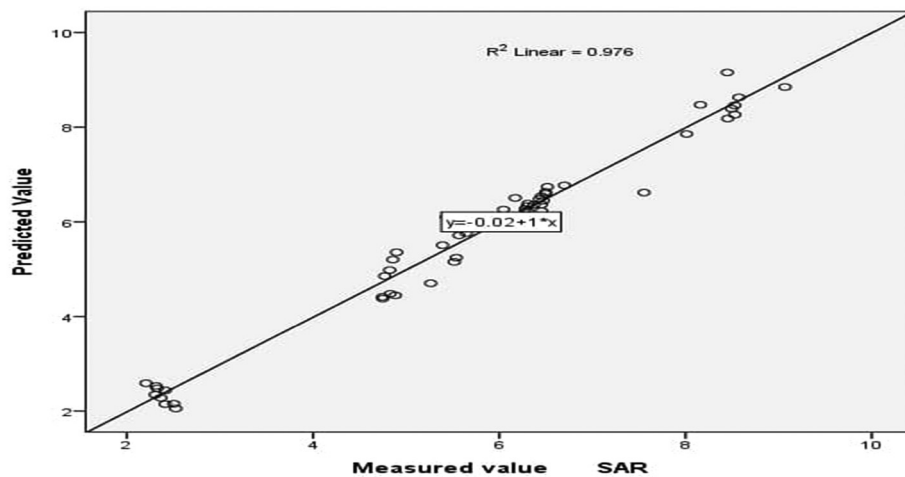
The applicability of ANNs was examined to estimate the salinity of groundwater. In ANNs, the choosing of input elements is essential to staging, as it impacts the performance of the prediction process. Variables such as EC, Ca^{2+} , Mg^{2+} , Na^+ , K^+ , and pH were adopted as input elements to the estimation of three salinity indices that



(a)



(b)



(c)

Figure 5: The comparison between measured and estimated salinity indices. (a) The comparison for TDS, (b) the comparison for Na%, and (c) the comparison for SAR.

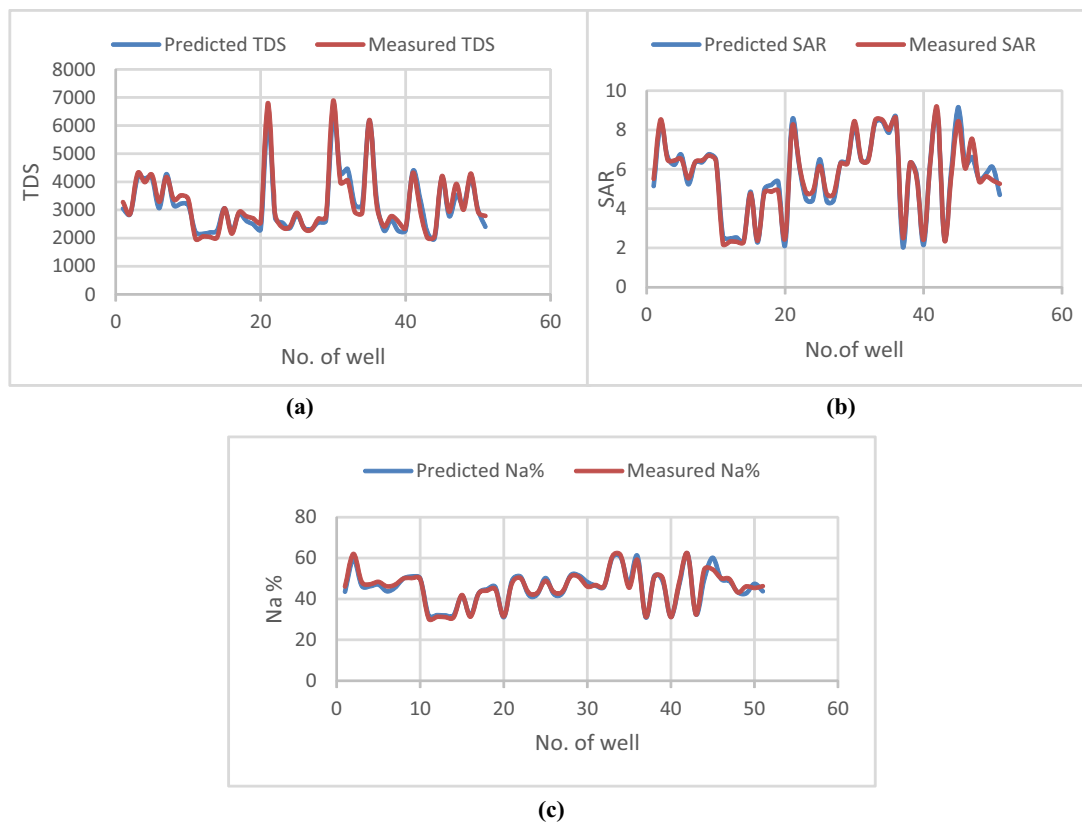


Figure 6: The variations of predicted and measured values. (a) The values of TDS between predicted and measured. (b) The values of SAR between predicted and measured. (c) The values of Na percent between predicted and measured.

characterizes the groundwater. These indices are TDSs, SAR, and Na percent.

The statistical outputs of the ANNs are presented in Table 3. In training and testing steps, the relative error and R^2 values of the estimated TDS were 0.038 and 0.95, SAR were 0.014, and 0.96, Na percent were 0.012 and 0.96, respectively. While these values predicted in testing step for TDS were 0.053 and 0.96, SAR were 0.065 and 0.97, and Na percent were 0.0113 and 0.96, respectively.

The results shown in Table 3 indicate that the performance of the selected ANN model is precise in general. The precision of the proposed model is due to the high R^2 and low relative error values. Also, it can be seen that the four neurons are giving minor error for estimation salinity indices of the groundwater. Figure 5 shows the comparison plots between measured and estimated salinity indices. It can be noticed that the model in the stage of testing is closer, in general, to the agreement fit line.

The variations of predicted and measured data were also shown in Figure 6. As shown in the figure, the modeled data tracked the measured data closely. Therefore,

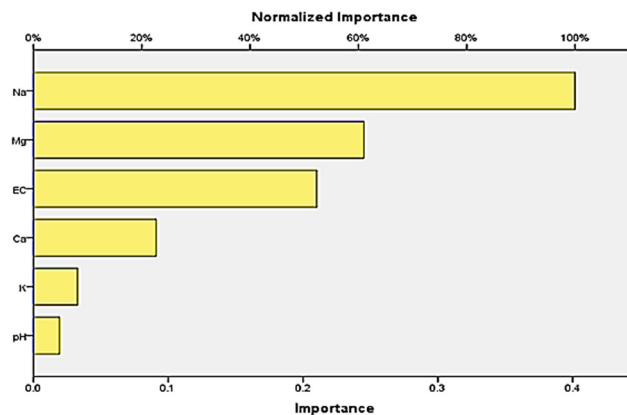


Figure 7: Input parameter importance.

by using simply measurable elements as input elements, ANNs can predict the TDS, SAR, and Na percent in groundwater, precisely.

The comparison between estimated and measured salinity indices of TDS, SAR, and Na percent based on linearity and R^2 shows the reliability of the assessed ANN model.

6 Parameter sensitivity analysis

The significance of the input parameters (independent variables) lies in the measurement of the sensitivity of the model through the estimation of the output elements. In this context, Figure 7 presents that Na exerts the highest effect on the method adopted by the ANN model for the salinity prediction, followed by Mg, EC, Ca, K, and pH.

7 Conclusion

This study deals with predicting groundwater salinity in the plateau of Najaf–Kerbala provinces using the SPSS V24 software to build the ANN model. The salinity of groundwater in terms of TDS, SAR, and Na% was predicted by using K^+ , EC, Ca^{2+} , Mg^{2+} , Na^+ , and pH as input elements to the model. The proposed ANN model with a structure of 6-4-3 presented a suitable ability to achieve the association between output and input elements. In the testing stage, the predicted output of TDS, SAR, and Na percent followed the measured data with an R^2 of 0.96, 0.97, and 0.96 with relative error of 0.038, 0.014, and 0.021, respectively. While in the training stage, the value of R^2 for TDS, SAR, and Na percent is 0.95, 0.96, and 0.96, respectively, with relative error of 0.053, 0.065, and 0.133, respectively. The high R^2 and smallest values of error estimations for training and testing samples indicate that the designed ANN network was acceptable. The Na, Mg, and EC, the most significant elements, impacted the model estimation. The proposed model could be utilized for entering new inputs to predict the groundwater salinity for irrigation purposes at the Najaf–Kerbala plateau in Iraq.

Conflict of interest: Authors state no conflict of interest.

References

- [1] Heidarzadeh N. A practical low-cost model for prediction of the groundwater quality using artificial neural networks Nima Heidarzadeh. *J Water Supply Res Technol - AQUA*. 2017;66(2):86–95. doi: 10.2166/aqua.2017.035.
- [2] Ghalib HB, Al-Hawash AB, Al-Qurnaw WS, Hameed Sultan B, Al-Enzy AW. Marshes waters sources hydrochemistry of the Bahr Al-Najaf at Najaf Province, Iraq. *J Phys Conf Ser*. 2019;1279(1):012059. doi: 10.1088/1742-6596/1279/1/012059.
- [3] Banerjee P, Singh VS, Chattopadhyay K, Chandra PC, Singh B. Artificial neural network model as a potential alternative for groundwater salinity forecasting. *J Hydrol*. 2011;398(3–4): 212–20. doi: 10.1016/j.jhydrol.2010.12.016.
- [4] Khudair BH, Jasim MM, Alsaqqar AS. Artificial neural network model for the prediction of groundwater quality. *Civ Eng J*. 2018;4(12):2959–70. doi: 10.28991/cej-03091212?.
- [5] Kheradpisheh Z, Talebi A, Rafati L, Ghaneian MT, Ehrampoush MH. Groundwater quality assessment using artificial neural network: a case study of Bahabad plain, Yazd, Iran. *Desert*. 2015;20(1):65–71. doi: 10.22059/jdesert.2015.54084.
- [6] Kılçaslan Y, Tuna G, Gezer G, Gulez K, Arkoc O, Potirakis SM. ANN-based estimation of groundwater quality using a wireless water quality network. *Int J Distrib Sens Netw*. 2014;2014:1–8. doi: 10.1155/2014/458329.
- [7] Sunayana K, Kalawapudi O, Dube Sharma R. Use of neural networks and spatial interpolation to predict groundwater quality. *Env Dev Sustain*. 2020;22(4):2801–16. doi: 10.1007/s10668-019-00319-2.
- [8] Aryafar A, Khosravi V, Zarepourfard H, Rooki R. Evolving genetic programming and other AI-based models for estimating groundwater quality parameters of the Khezri plain, Eastern Iran. *Environ Earth Sci*. 2019;78(3):69. doi: 10.1007/s12665-019-8092-8.
- [9] Maroufpoor E, Jalali S, Nikmehr M, Shiri S, Shiri N, Maroufpoor J. Modeling groundwater quality by using hybrid intelligent and geostatistical methods. *Env Sci Pollut Res*. 2020;27(22):28183–97.
- [10] Bedi S, Samal A, Ray C, Snow D. Comparative evaluation of machine learning models for groundwater quality assessment. *Environ Monit Assess*. 2020;192(12):776. doi: 10.1007/s10661-020-08695-3.
- [11] Al-kubaisi QY, Al-abadi AM, Al-ghanimy MA. Mapping groundwater quality Index for irrigation in the Dibdibba aquifer at Karbala - Najaf plateau, central of Iraq. *Iraqi J Sci*. 2018;59(3C):1636–52. doi: 10.24996/ij.s.2018.59.3c.10.
- [12] Lallahem S, Hani A. Artificial neural networks for defining the water quality determinants of groundwater abstraction in coastal aquifer. *AIP Conf Proc*. 2017;1814(1):020013. doi: 10.1063/1.4976232.
- [13] Wagh VM, Panaskar DB, Muley AA, Mukate SV, Lolage YP, Aamalawar ML. Prediction of groundwater suitability for irrigation using artificial neural network model: a case study of Nanded tehsil, Maharashtra, India. *Model Earth Syst Env*. 2016;2(4):1–10. doi: 10.1007/s40808-016-0250-3.
- [14] Nasr M, Tawfik A, Ookawara S, Suzuki M. Prediction of hydrogen production using artificial neural network. Conference: Seventeenth International Water Technology Conference, IWTC17, Istanbul; 2013. p. 5–7, no. November.
- [15] Asghari Moghaddam A, Nadiri, Fijani E. Ability to study different Models of artificial neural networks to evaluate groundwater water level in the hard formation; 2006.