Advances in Environmental Technology 11(4) 2025, 396-413.

Iranian Reasearch Organization for Science and Technology (IROST)

Advances in Environmental Technology

Journal home page: https://aet.irost.ir

Water's corrosion and scaling potential prediction using artificial neural networks and gene expression programming in several rural water distribution networks in Kermanshah Province, Iran

Akram Fatemi[a]*, Shabnam Vaisi[a], Rasool Ghobadian[b]

[a] *Department of Soil Science, Razi University, Kermanshah, Iran.*

[b] *Department of Water Science and Engineering, Razi University, Kermanshah, Iran.*

ARTICLE INFO

Document Type:
Research Paper

Article history:
Received 27 October 2024
Received in revised form 23 June 2025
Accepted 01 July 2025

Keywords:
Aggressive index (AI)
Langelier saturation index (LSI)
Larson–Skold index (L-SI)
Puckorius scaling index (PSI)
Ryznar stability index (RSI)

ABSTRACT

Water quality causes severe restrictions on the utilization of water resources. Corrosion and scaling are the most common problems in the operation and maintenance of water facilities. Corrosive indices are an indirect method of detecting and measuring water's tendency to corrosion and scaling. Water corrosion and scaling are complex phenomena that cannot be easily modeled. This study used meta-heuristics methods, such as artificial neural networks (ANN) and gene expression programming (GEP), to predict the water's corrosion and scaling potential of the distribution network in some rural areas of Kermanshah Province. Equations were extracted to estimate water corrosion and scaling indices using linear regression and GEP. The results showed that ANN could reveal water corrosion and scaling indices with the highest correlation coefficient (0.95, 0.91, 0.96, 0.92, and 0.99) and the lowest percentage errors (0.20, 0.44, 0.40, 0.44, and 0.08) for the Langelier saturation index (LSI), Ryznar stability index (RSI), Puckorius scaling index (PSI), Aggressive index (AI), and Larson–Skold index (L-SI), respectively. Also, the linear and nonlinear relationships obtained by a high-precision GEP model (0.80 to 0.97) can estimate corrosion and scaling indices with lower cost and more accuracy by measuring the most influential physicochemical parameters.

1. Introduction

Water quality is affected by physical, chemical, and biological changes due to human and natural factors in any region. Most of these changes are harmful and cause severe restrictions on the use of water resources. One of the most common problems in the operation and maintenance of water facilities, especially groundwater resources, is corrosion and scaling [1]. Corrosion is the primary cause of chemical property alteration, efficiency loss, and the reduction of the life span of water infrastructure. In addition, the substantial financial burden of water infrastructure due to corrosion, health, and safety costs for society

*Corresponding author Tel.: +98 9183894008
E-mail: a.fatemi@razi.ac.ir
DOI: 10.22104/AET.2025.7194.1980

should also be considered [2]. Currently, a significant percentage of the annual income of countries is allocated to corrosion and scaling issues. In the US more than $300 billion is spent annually on the corrosion, maintenance, or replacement of products [3, 4].

The study of urban treated water loss in Iran shows that more than 30% of distributed water is wasted annually due to decay caused by corrosion of water distribution pipes [5, 6]. The tendency of water's corrosivity or scaling is determined by examining water stability. Stable water tends to have low corrosiveness and scaling; its values vary for use type [7]. Water corrosion and scaling indices indirectly detect and measure the water's tendency for corrosion and scaling. LSI, RSI, AI, PSI, and L-SI are the most common corrosion and scaling indices [8]. Using several of these corrosion and scaling indices simultaneously can provide a more confident water equilibrium status for control measures [9].

This section explores research on the utilization of indices for assessing the potential of water's corrosion and scaling. Gholizadeh et al. [10] evaluated the scaling and corrosion potential of groundwater resources in the Yazd-Ardakan Plain, Iran. Nayeria et al. [11] evaluated the scaling and corrosion potential of drinking water in the distribution network of Kermanshah City, Iran, during the 2018 winter and summer seasons. Fatemi et al. [12] studied the scaling and corrosion potential of drinking water in particular rural distribution networks in different climate zones of Kermanshah Province, Iran, from 2009 to 2017. Siddha and Sahu [13] evaluated the groundwater of Central Gujarat for industrial usage according to LSI, PSI, RSI, L-SI, corrosivity ratio (CR), Revelle index (RI), and chloride sulfate mass ratio (CSMR). Estimating the water's corrosion and scaling potential requires measuring many physicochemical parameters of water, which is costly and time-consuming. However, the complex phenomena of water's corrosion and scaling cannot be easily modeled due to various factors causing this complexity. The physical, chemical, and electrochemical reactions and processes are essential to predict and model corrosion. The modeling is based on these factors. Despite the mechanistic models' successes, some influential factors are unknown. Therefore, models that more accurately predict and model the potential of corrosion are needed [14].

Some researchers have examined the high ability of machine learning models to model complex and nonlinear systems in hydrology and water resources management-related problems across various regions [15, 16]. The studies have reported an accurate prediction of some water quality parameters using machine learning models. The use of the ANN model, with the increasing development of computational methods such as artificial intelligence, has been widely used in studies on the prediction of different parameters of water resources [17-19]. Researchers have emphasized the high accuracy of this method compared to experimental and regression relationships [20-23]. ANN is more valid because it uses proven formulas with minimal errors [24]. The GEP model is based on an evolutionary algorithm, used to estimate nonlinear phenomena based on mathematical relationships using a tree structure [25, 26]. So far, no study has been conducted to predict the water's corrosion and scaling potential using ANN, GEP, or other machine-learning models.

This study aimed to provide advanced and up-to-date models to predict the corrosion and scaling potential of water in distribution networks in some rural areas of Kermanshah Province. ANN and GEP modeling were explored to identify the most influential physicochemical parameters of water. By achieving this goal, instead of measuring all the involved physicochemical parameters of water, it would be possible to accurately predict the corrosion and scaling potential of drinking water using fewer parameters. Consequently, time and costs would be reduced. Another objective was to develop equations that are more precise and feasible than empirical indices (LSI, RIS, AI, PSI, and L-SI) for predicting the corrosion and scaling potential of drinking water.

2. Materials and Methods

2.1. Study area

Kermanshah Province is located at 33° 40′ and 35° 18′ N latitude and 45° 24′ and 48° 07′ E longitude. This region has an area of about 25045 $km^{2,}$ with six types of climatic regions: cold semi-dry, warm semi-dry, warm and dry, cold Mediterranean, moderate humid, and semi-humid climate zones (based on revised Embereger) (Figure 1). According to long-term meteorological data, the average annual temperature in the different climate zones ranges from 12.8 to 22.0°C, and the annual rainfall is about 290–756 mm.

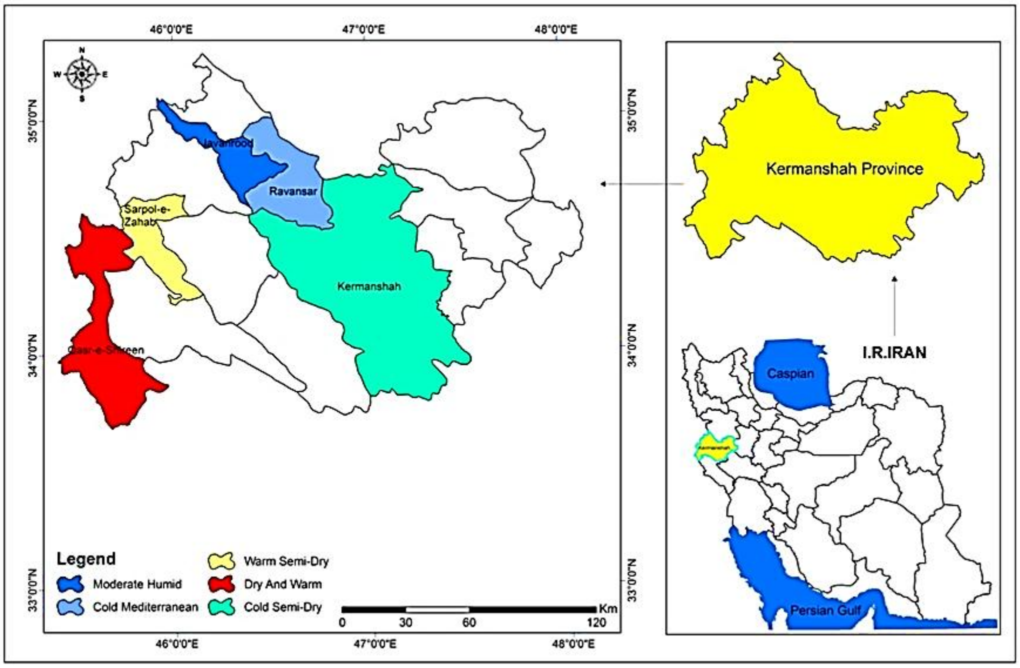

Fig. 1. Location of the study area in Kermanshah Province, Iran.

2.2. Datasets

Considering the different climatic zones of Kermanshah Province, 32 sampling datasets from stations located in five cities were obtained from the city's hydraulic works. The datasets included results for corrosion characteristics that were monitored from 2009 to 2017: alkalinity, pH, calcium hardness, temperature (T), total dissolved solids (TDS) [27]. The presence of anions chlorine (Cl^-), carbonate (CO_3^{2-}), bicarbonate (HCO_3^-), and sulfate (SO_4^{2-}) were also included. The monitoring stations in Javanrood, Ravansar, Kermanshah, Sarpol-e-Zahab, and Qasr-e-Shireen contributed to moderate humid, cold Mediterranean, cold semi-dry, warm semi-dry, and dry and warm climate zones, respectively (Table 1).

2.3 Water's corrosion and scaling indices

The LSI, RSI, PSI, AI, and L-SI indices were calculated using Equations 1 to 5 [8, 27, 28].

$$\mathrm{LSI} = \mathrm{pH} - \mathrm{pH_s} \tag{1}$$

$$\mathrm{RSI} = 2\mathrm{pH_s} - \mathrm{pH} \tag{2}$$

$$\mathrm{PSI} = 2\mathrm{pH_S} - \mathrm{pH_{eq}} \tag{3}$$

$$\mathrm{AI} = \mathrm{pH} + \log(\mathrm{Alk} \times \mathrm{H})\mathrm{AI} = \mathrm{pH} + \log(\mathrm{Alk} \times \mathrm{H}) \tag{4}$$

$$\mathrm{L-SI} = \frac{(\mathrm{SO_4}^{2-} + \mathrm{Cl}^-)}{(\mathrm{HCO_3}^- + \mathrm{CO_3}^{2-})} \tag{5}$$

In these equations, *pH_s* is the *pH* of water saturated with calcium carbonate, which was calculated using Equation 6:

$$\mathrm{pH_s} = (9.3 + \mathrm{A} + \mathrm{B}) - (\mathrm{C} + \mathrm{D}) \tag{6}$$

$$\begin{aligned}
&\mathrm{A} = (\log\ [\mathrm{TDS}] - 1)/10 \\
&\mathrm{B} = -13.12 \times \log[\mathrm{T} + 273] + 34.55 \\
&\mathrm{C} = \log[\mathrm{H}] - 0.4 \\
&\mathrm{D} = \log[\mathrm{T.Alk}] \\
&\mathrm{Alk} = \mathrm{T.Alk}(\mathrm{mg\ L^{-1} as\ CaCO_3})
\end{aligned}$$

In these equations, TDS is the total dissolved solids in terms of mg L^{-1}, T, the temperature in °C, Ca^{2+} as $CaCO_3$, H, calcium hardness, and T.Alk, total alkalinity in terms of calcium carbonate (mg L^{-1}). To calculate pH_{eq} (water saturation pH), Equation 7 was used:

$$\mathrm{pH_{eq}} = 1.465 \times \log(\mathrm{T.Alk}) + 4.54 \tag{7}$$

2.4. ANN

The ANN is one of the subcategories of artificial intelligence for information processing inspired by the brain's biological nervous system. It processes information, in which millions of neurons may solve problems or store information by communicating with each other. The learning process in ANN is done through training using input and output data. A set of correct inputs and outputs is given to the network, and the ANN uses these inputs to create a complex mathematical model that produces the proper response in case of new inputs [29]. Neural networks consist of layers and are usually created in a regular form. Input information and data are entered into the first layer, which is the input layer. The middle, hidden, and last layers that provide the model output solutions are the output layers [30]. Many functions can be used to introduce non-linearity in neural networks and also to create training data of the artificial neural networks, called transfer functions. These functions include Sigmoid, Gaussian, Hyperbolic tangents, and Hyperbolic secant [31].

In this study, the effect of physicochemical parameters of water on water's corrosion and scaling indices was investigated using an ANN. From the 146 series of information available for each parameter, 90 sets were used for training the model, 36 sets were used for validation, and 20 groups were used for the final test of the model. Also, the Sigmoid function was selected as the transfer function to implement the model. This study used Qnet2000 software [32] to evaluate an ANN model for estimating the water's corrosion and scaling indices. Neural networks used for the following indices had the lowest error percentage. The LSI and RSI indices' six input parameters were pH, TDS, T, Ca^{2+}, CO_3^{2-}, and HCO_3^- with five nodes in the hidden layer. The schematic representation of the ANN for LSI is shown in Figure 2.

Five input parameters were chosen for PSI (TDS, T, Ca^{2+}, CO_3^{2-}, and HCO_3^-) and AI (pH, CO_3^{2-}, HCO_3^-, Ca^{2+}, and Mg^{2+}) with four nodes in the hidden layer. Four input parameters, including Cl^-, SO_4^{2-}, CO_3^{2-}, and HCO_3^- with three nodes in the hidden layer, were selected for L-SI. Then, the ANN model was implemented, and the simulation was carried out up to 10,000 times. These parameters were chosen to examine the influence of each parameter on the individual corrosion index. For this purpose, the contribution of each physicochemical parameter of water was evaluated using sensitivity analysis. The output of this part would help to develop new models to estimate the water's corrosion and scaling potential. Afterward, linear relationships [30] between water corrosion and scaling indices and physicochemical parameters of water were investigated using SPSS 20.0 software (SPSS Inc., Chicago, Illinois, USA).

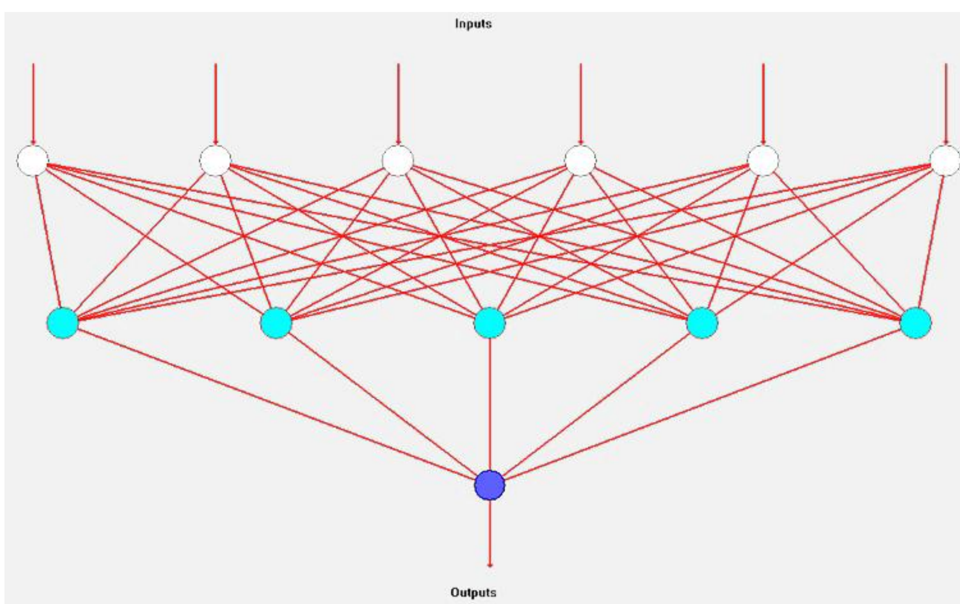

Fig. 2. schematic representation of an artificial neural network for LSI

Table 1. The average amounts of physicochemical parameters of water in distribution networks in different climate zones of Kermanshah Province

Climate zones		EC	Ca^{2+}	CO_3^{2-}	HCO_3^-	Cl^-	SO_4^{2-}	pH	T*	TDS**	H***	T.Alk****
		μmhos cm^{-1}	mg L^{-1}						^{0}C	mg L^{-1}		
Moderate humid	Min	305.00	49.70	0.01	1.89	0.06	0.04	7.30	16.20	198.40	174.00	116.00
	Max	435.00	75.50	0.23	5.23	0.21	0.21	7.50	27.80	269.70	252.00	320.00
	Mean	362.40	61.34	0.01	3.20	0.10	0.08	7.55	22.08	221.43	199.90	196.50
	SD	43.51	8.86	0.01	0.67	0.03	0.03	0.15	4.12	26.74	24.93	41.01
Cold Mediterranean	Min	350.00	47.50	0.01	2.64	0.07	0.03	7.40	13.50	217.00	174.00	150.00
	Max	541.00	79.60	0.03	4.05	0.61	2.80	8.00	27.90	432.10	304.00	248.00
	Mean	429.25	61.88	0.01	3.43	0.27	0.15	7.65	22.59	268.11	222.00	210.97
	SD	71.29	8.12	0.01	0.36	0.27	0.38	0.18	4.35	40.04	25.45	22.09
Cold semi-dry	Min	348.00	41.80	0.00	2.96	0.01	0.05	7.00	17.00	220.00	168.00	178.00
	Max	371.00	65.70	0.05	4.08	0.94	0.16	8.30	25.40	228.74	204.00	206.00
	Mean	319.00	46.47	0.02	3.27	0.15	0.13	7.70	22.62	228.68	202.90	199.48
	SD	115.00	4.53	0.01	0.25	0.16	0.03	0.29	2.04	25.98	25.43	15.26
Warm semi-dry	Min	499.00	20.10	0.00	2.20	0.15	0.21	7.20	12.70	288.30	208.00	140.00
	Max	712.00	121.81	0.14	6.09	0.71	0.61	8.10	27.90	441.30	384.00	323.00
	Mean	588.97	67.65	0.01	4.34	0.27	0.33	7.59	21.24	368.19	305.10	271.97
	SD	55.78	28.61	0.02	1.04	0.13	0.12	0.18	4.52	39.75	40.18	40.25
Dry and warm	Min	467.00	55.30	0.01	3.50	0.10	0.15	7.60	12.50	289.40	238.00	220.00
	Max	483.00	68.10	0.03	4.60	0.61	0.22	7.80	27.10	305.40	282.00	284.00
	Mean	476.16	64.86	0.02	4.02	0.17	0.19	7.70	19.28	295.63	264.83	247.16
	SD	5.44	4.37	0.01	0.32	0.14	0.02	0.09	5.77	4.40	16.36	19.73

* Temperature; **Total Dissolved Solids; *** Hardness as $CaCO_3$; **** Total Alkalinity as $CaCO_3$

2.5. GEP

The GEP method, introduced by Ferreira in 1999, is a meta-heuristic optimization algorithm to achieve a correct estimate of the goal variable. It is based on circulatory algorithms using Darwin's evolutionary theory [33]. This method combines Genetic Programming (GP) and Genetic Algorithm (GA). This method combines linear and simple chromosomes with the same fixed length with GA and branched structures of different sizes and shapes, similar to decomposition trees in GP. In GEP, improvements occur in a linear structure and are then expressed as a tree structure, causing only the modified genome to be transferred to the next generation. Therefore, there is no need to replicate and mutate rather cumbersome structures [34]. In GEP, the information of chromosomes is decoded in a tree mode called translation. According to this process, chromosomes are called from left to right and from top to bottom [35]. Figure 3 represents the flowchart of a gene expression algorithm [36]. An example of translating a chromosome into a tree-shaped state is also shown in Figure 4 [37].

The GEP method and GeneXprotools 4.0 software [38] were used to investigate the nonlinear relationships between water's corrosion and scaling indices and physicochemical parameters of water. The GEP method was performed with two genes, the root relative squared error (RRSE) fitting function, 95 educational samples, and 30 samples as a test, with a mutation rate of 0.066 and an inversion rate of 0.1.

2.6. Statistical Indices

Various evaluation criteria, i.e., correlation coefficient (*r*), maximum error (Max Error), mean square error (*MSE*), standard deviation (*STD*), root mean square error (*RMSE*), and mean absolute error (*MAE*) were used to assess the performance of models (Equations of 8-13):

$$r = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\Sigma(x_i - \bar{x})^2 \Sigma(y_i - \bar{y})^2}} \quad (8)$$

where *r* is the correlation coefficient x_i, and y_i, are the values of the *x* and *y* variables, *and* $\bar{x}$ and $\bar{y}$ are the mean of the values of the *x* and *y* variables.

$$\text{Max Error} = \frac{(\hat{x}_i - x_i)}{\hat{x}_i} \quad (9)$$

where *Max error* is the highest percentage of errors.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (10)$$

where *MSE* is the mean square error, y_i is the measured observation, $\hat{y}_i$is the predicted observation, and *n* is the number of observations.

$$STD = \sqrt{\frac{\Sigma(x_i - \bar{x}_i)^2}{N}} \quad (11)$$

STD is a standard deviation that shows the distance of each observation (*xi*) from the average of observations ($\bar{x}i$), and *N* is the number of observations.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \hat{x}_i)^2}{N}} \quad (12)$$

In this equation, *RMSE* is the root mean square error, x_i is the measured observation, $\hat{x_i}$ is the actual observation, and *N* is the number of observations.

$$MAE = \frac{\sum_{i=1}^{n}|y_i - x_i|}{n} \quad (13)$$

where *MAE* is the mean absolute error, y_i is prediction, x_i is true value, and *n* is the total number of data points. Generally, a robust model is characterized by a high R^2 value and low error values. This means that the model can accurately reflect the essential relationships in the data and generate reliable predictions.

3. Results and Discussion

3.1. Water's corrosion and scaling potential indices

The water's corrosion and scaling potential indices, including LSI, RIS, AI, PSI, and L-SI, were considered. LSI can qualitatively assess water potential in calcium carbonate deposition formation [39]. This index shows the effect of Ca^{2+}, total alkalinity (T.ALK), TDS, and T in calculating saturated pH values. In other words, this index can be mentioned as pH changes to reach the water equilibrium [1].

RSI is a modified type of LSI whose values are positive against LSI. This index quantitatively indicates the dependence between calcium carbonate saturation and crust formation. The determinants of this index are specified by actual pH, the concentration of Ca^{2+} ions, HCO_3^-, TDS, and T [40]. LSI and RSI show the difference between the actual pH of the water and the pH saturated by calcium carbonate [41]. The PSI, also known as the practical scaling index, shows the water's corrosion and scaling potential. Unlike the other three indices, it is not associated with the actual pH of water. PSI values are affected by water buffering capacity, and equilibrium pH is used instead of actual water pH to calculate its effect [42, 43]. The AI is effective for asbestos cement pipes and for a temperature range of 4 to 27°C. The AI is mainly a function of pH, Ca^{2+} concentration, and alkalinity [44]. The L-SI shows the impact of SO_4^{2-}, Cl^-, CO_3^{2-}, and HCO_3^- anions. CO_3^{2-} and HCO_3^- reduce the water's corrosion, while Cl^- and SO_4^{2-} increase it [45].

The water's corrosion and scaling status based on the LSI is divided into three categories: scaling (LSI>0), neutral (LSI> 0.02), and corrosive (LSI<0). This classification for the RSI is as follows: scaling (RSI<5), neutral (5<RSI<7), and corrosive (RSI>7); as well, PSI has two forms of scaling: (PSI<6) and corrosive (PSI>6)[43]. The AI has low corrosion (AI>12), moderate corrosion (10<AI<12), and severe corrosion (AI<10). For the L-SI, the amounts (L-SI< 0.8) of film formation (0.8<L-SI<2.1), low corrosion, and (L-SI> 1.2) high corrosion are considered [8, 27, 46].

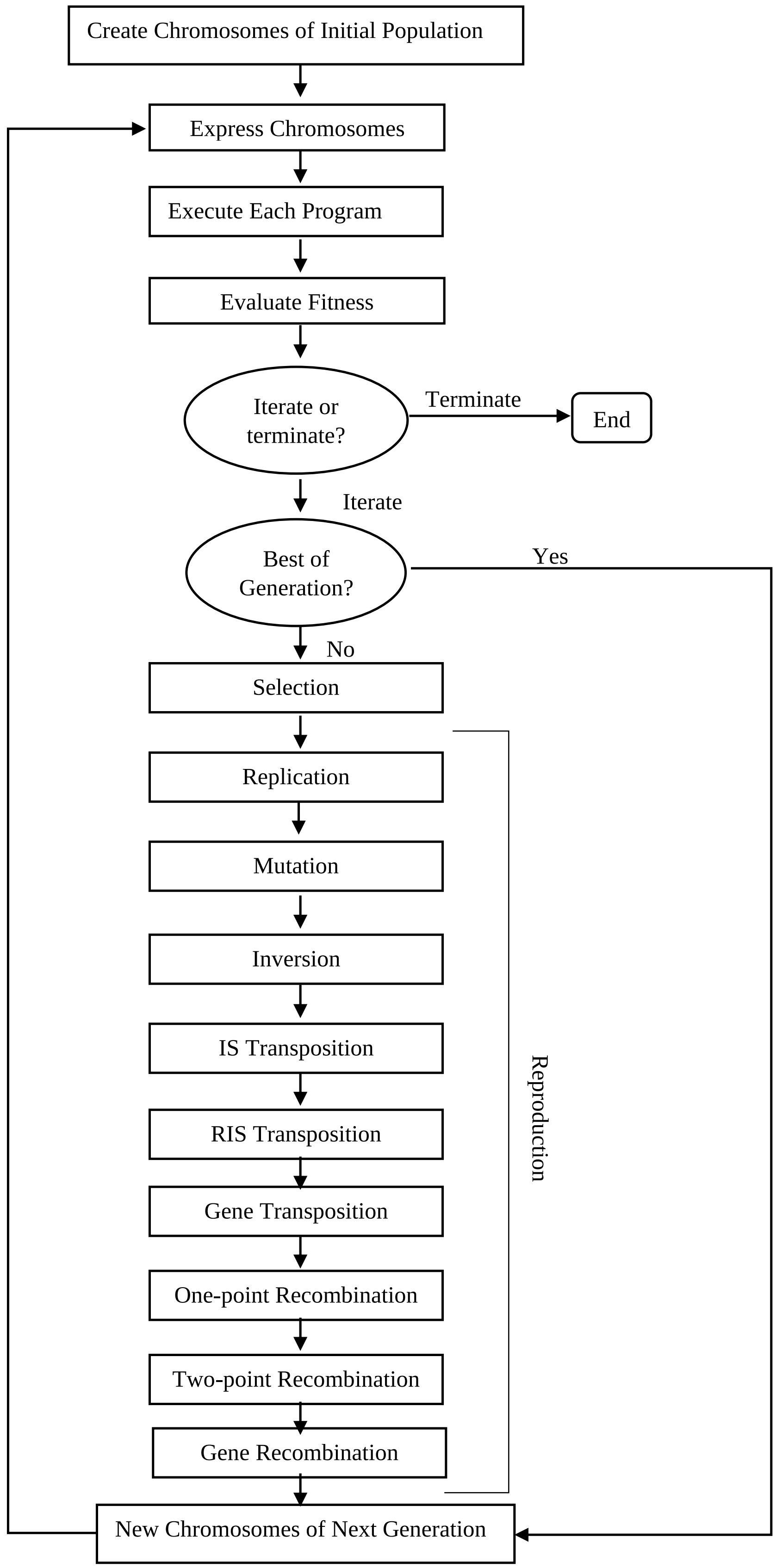

Fig. 3. The flowchart of a gene expression algorithm [36]

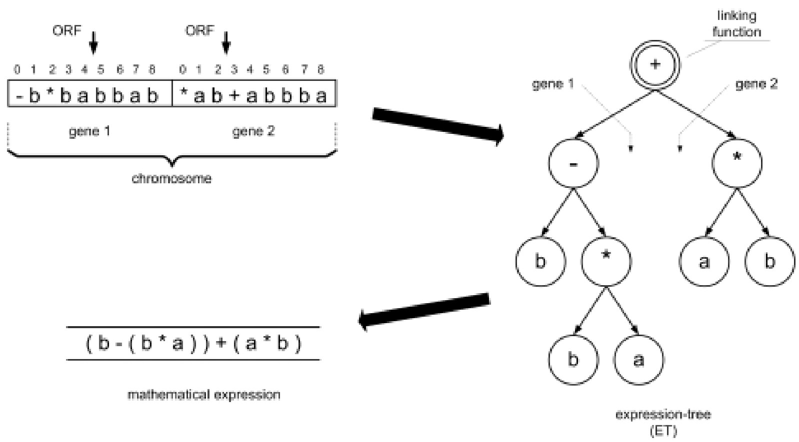

Fig. 4. presentation of GEP chromosomes [37].

The results of studying the water's corrosion and scaling potential in different regions have shown that water is corrosive in many areas [47]. For instance, according to RSI and PSI, Gholizadeh et al. [10] reported low to moderate corrosivity of most groundwater resources in the Yazd-Ardakan Plain in Iran. On the other hand, the LSI and AI classified water as having a minor tendency to generate scale and as non-aggressive, respectively. They concluded that RSI and LSI, especially RSI, were more valuable than PSI and AI for the studied area. Taghavi et al. [48] assessed the corrosivity of Iranshahr's drinking water resources, reporting LSI values between -1.53 and -0.96, RSI from 9.63 to 10.54, PSI from 9.05 to 10.68, and AI between 12.04 and 12.91. These findings, coupled with those of Amouei et al. [49], who reported AI means of 6.17 and 6.27, indicated a high percentage of corrosive water samples (82.2%–100%) based on various indices.

Siddha and Sahu [13] evaluated the groundwater of Central Gujarat for industrial usage according to LSI, PSI, RSI, L-SI, corrosivity ratio (CR), Revelle index (RI), and chloride sulfate mass ratio (CSMR). According to LSI and PSI, most of the analyzed area's groundwater tended to scale in water distribution systems.

Nayeria et al. [11] evaluated the scaling and corrosion potential of drinking water in the distribution network of Kermanshah City during 2018 in the winter and summer seasons. According to their results, the drinking water distribution network tended to be corrosive. Fatemi et al. [12] studied the scaling and corrosion potential of drinking water in the particular rural distribution networks in different climate zones of Kermanshah Province from 2009 to 2017. Based on water characteristics in various climate zones, LSI and RSI were chosen as good indices for the water's corrosion and scaling potential indices in different climate zones [12]. Gholizadeh et al. [10] reported the same results for groundwater resources in the Yazd-Ardakan Plain, Iran. They concluded that RSI and LSI, especially RSI, were more valuable than PSI and AI for the studied area. Siddha and Sahu [13] evaluated the groundwater of Central Gujarat for industrial usage according to LSI, PSI, RSI, L-SI, corrosivity ratio (CR), Revelle index (RI), and chloride sulfate mass ratio (CSMR). Based on LSI and PSI, the groundwater of a major portion of the studied area tended to be scaling in water distribution systems.

3.2. ANN

The results of the ANN model for the water's corrosion and scaling indices are shown in Table 2. As indicated in Table 2, ANN produced highly satisfactory correlation coefficient (R^2) results. Also, a low level of maximum error in each subset and the MSE serve as a high level of confirmation of actual and predicted data. R^2 values were generally over 0.94 and ranged from 0.88 to 0.99 within subsets (training, test, and recall). Additionally, the maximum percentage of error values for all the water's corrosion and scaling indices did not exceed 4% (on average). The lowest MSE values of 0.00 and 0.01 were yielded on testing and training data for L-SI, respectively. These findings are corroborated by results from previous studies [22, 23, 50, 51]. The results of Kulisz and Kujawska [22] indicated a high level of conformity (R^2 (general and within subsets) 0.98, MSE in each subset (training, test, validation), along with RMSE at a level of 0.62) of the actual data of Warta River's WQI compared to those obtained during prediction by ANN.

To gain more information on ANN and GEP applications for predicting water's corrosion and scaling potential, the literature regarding the use of these or other models in predicting water quality was reviewed. The results of the study [52] showed the high ability of neural networks to predict the underground water quality of the Birjand Plain by accurately estimating nitrate-sodium (R^2=0.96), magnesium (R^2=0.997), calcium (R^2=0.93), and sodium (R^2=0.98) concentrations. Moreover, it was revealed that ANN precisely estimated some water's physico-chemical parameters which reviewed in the following. TDS of aquifers in the Tehran Plain (R^2=0.96, NRMSE=0.175) [53]; EC, sodium absorption ratio (SAR), and TDS of the Karoun River [54]; and groundwater salinity ($R^2_{training}$= 0.64, $R^2_{validation}$=0.67 and R^2_{test}=0.90) [4], EC, SAR, and TDS of underground water resources of the Mehran and Dehloran Plains [55] were predicted successfully by ANN.

They reported that the coefficient of determination of EC, SAR, and TDS for training, validation, and testing was over 90 percent. Mirzavand et al. [56] simulated (with R^2 values over 90 percent) the quality parameters of groundwater resources accurately in the Kashan Plain, which included the water table level, annual rainfall height, Cl^- concentration in the previous year, and Cl^- concentration in the current year using ANN. The results showed the high accuracy of the ANN model in prediction and simulation [56].

Table 2. The results of modeling water's corrosion and scaling indices using an ANN model.

Index	Data	R^{2*}	Max Error**	MSE***	STD****
LSI	Training Data	0.95	0.20	0.04	0.07
	Test Data	0.94	0.23	0.04	0.08
	RECALL	0.96	0.18	__	0.08
RSI	Training Data	0.91	0.44	0.05	0.16
	Test Data	0.88	0.40	0.06	0.16
	RECALL	0.96	0.23	__	0.08
PSI	Training Data	0.92	0.44	0.05	0.15
	Test Data	0.84	0.62	0.07	0.21
	RECALL	0.97	0.16	__	0.06
AI	Training Data	0.96	0.40	0.03	0.06
	Test Data	0.93	0.30	0.04	0.09
	RECALL	0.99	0.08	__	0.04
L-SI	Training Data	0.99	0.08	0.01	0.01
	Test Data	0.98	0.03	0.00	0.01
	RECALL	0.99	0.01	__	0.00

* Correlation Index ** Maximum error *** Mean square error ****Standard deviation

Alaie et al. [17] investigated the water quality of the Neyshabur Plain by ANN and a fuzzy neural inference system. This study used quantitative and qualitative data on flow rate, temperature, CO_3^{2-}, HCO_3^-, and concentration of Cl^-, SO_4^{2-}, calcium (Ca^{2+}), magnesium (Mg^{2+}), and sodium (Na^+) to estimate TDS. Mohammadi et al. [18] reported the high conformity between experimental and predicted data by the ANN model for fluoride (F^-) concentrations in groundwater resources in Khaf, Iran. In a study by Emami et al. [19], the water quality of the Jolfa Plain was examined by two methods, ANN and the Imperialist Competitive Algorithm (ICA). Kulisz and Kujawska [22] used ANN to predict the Warta River's surface water quality index (WQI) in 2014-2018. The high correlation coefficient of 0.9792, low MSE in subsets, and RMSE of 0.624507639 confirmed the accuracy of the model. The error for WQI values was below 4%, indicating a high level of conformity between real and predicted data.

3.3. Sensitivity analysis

The sensitivity analysis results of the physicochemical parameters of water input to the model for each corrosion and scaling index of water are shown in Figure 5. The influence of the pH, TDS, T, Ca^{2+}, CO_3^{2-} and HCO_3^- parameters were 64.16, 2.61, 22.99, 0.85, 0.77, and 8.62 percent, respectively, for the LSI (Figure 5). For the RSI, the highest participation and influence were related to pH parameters with a value of 49.69, and the lowest was related to Ca^{2+} with 0.61%. For RSI, the percentage of influence of parameters TDS, T, CO_3, and HCO_3^- was 2.05, 36.7, 5.8, and 5.15, respectively (Table 3). The influence of the TDS, T, Ca^{2+}, CO_3^{2-}, and HCO_3^- parameters for the PSI was 7.05, 29.33, 6.9, 11.03, and 45.69%, respectively (Figure 5). The sensitivity analysis results on the AI showed that the influence of pH, CO_3^{2-}, HCO_3^-, Ca^{2+}, and Mg^{2+} parameters were 53.44, 1.45, 25.79, 8.65, and 10.72, respectively (Table 3). The sensitivity analysis results on the L-SI showed that the influence of the Cl^-, SO_4^{2-}, CO_3^{2-}, and HCO_3^- parameters were 32.22, 58.38, 9.06, and 0.34, respectively (Figure 5).

In general, the sensitivity analysis results demonstrated that pH was the most effective parameter, followed by temperature (T) for LSI and RSI. The pH and HCO_3^- were determined as more effective parameters than the others for AI. In addition, the influence of HCO_3^- was considerable for PSI and AI. However, TDS, CO_3^2, Ca^{2+}, and Mg^{2+} influences were generally less than the mentioned parameters for almost all the water's corrosion and scaling indices. The results of the study by Shah et al. [50] revealed that HCO_3^- was the most sensitive parameter influencing both TDS and EC models. Alqahtani et al. [23] reported that sensitivity analysis showed that HCO_3^- was the most effective variable, followed by Cl^- and SO_4^{2-} for both the EC and TDS predicted by the GEP, ANN, and random forest (RF) models. On the contrary, Aldrees et al. [21] reported that Cl^- and HCO_3^- had substantial impacts on the predictions of EC and TDS in the developed multi-expression programming (MEP) models. The water's most influential physicochemical parameters can be identified using machine learning models, thereby reducing costs and increasing prediction accuracy [57]. Hence, after determining the degree of influence of each of the physicochemical parameters of water, to determine equations that are capable of estimating the corrosion and scaling indices of water based on the simple models, several stages were introduced by eliminating the parameters with less influence.

The correlation coefficient (R^2) in the training stage, ANN test, and sensitivity analysis results for each of the models obtained are presented in Table 3. The R^2 coefficient values showed that the R^2 values between the models were relatively the same. On the other hand, removing parameters with less influence had no significant effect on the accuracy of the models presented for these indices. It helped to develop new equations with the least parameters, which are as accurate as empirical equations. The empty columns indicate the absence of the corresponding parameter in the studied index formula. Using these new equations of the water's corrosion and scaling indices with fewer physicochemical parameters, instead of all in empirical equations, helps to reduce costs.

Since R^2 values were obtained closely together (Table 3), linear relationships between corrosion and scaling indices and physicochemical parameters of water were investigated with the lowest number of effective parameters using the linear regression method. The results are reported in Table 4. The results also showed some significance and gave the lowest error, ranging from 0.06 to 0.24 in terms of root mean square error (MSE) for the equations (Table 4).

.

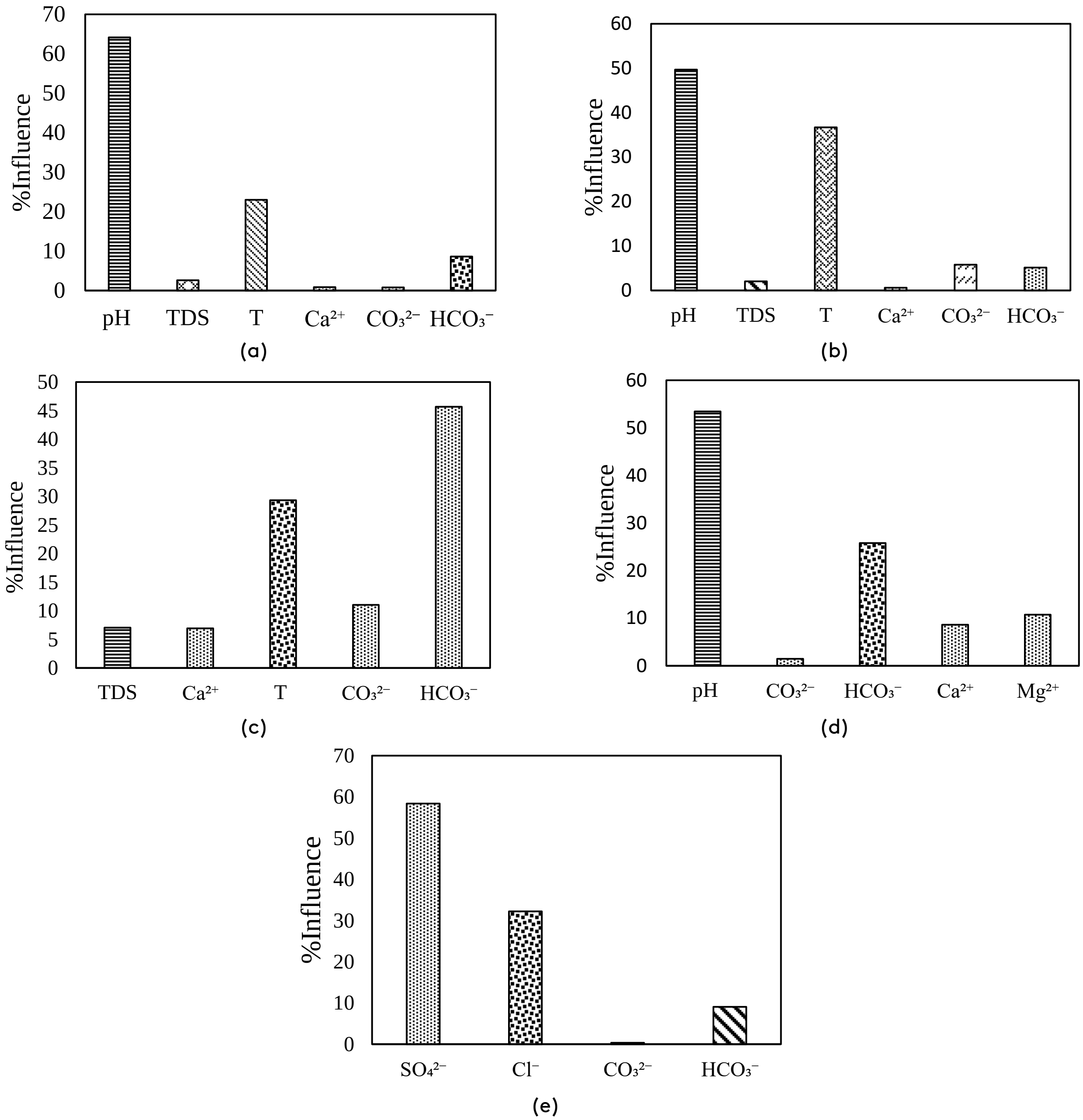

Fig. 5. Comparison of the influence of each of the influencing parameters on water corrosion and scaling indices: a) LSI, b) RSI, c) PSI, d) AI, and e) L-SI.

3.4. GEP

In addition to the extracted relationships through linear regression presented in Table 4, the GEP method (the tree model) was also extracted to predict the water's corrosion and scaling indices. The relationships between water's corrosion and scaling potential that were estimated by the GEP method, and the physicochemical parameters of water are reported in Table 5. The results indicated a strong correlation R^2 (0.84 to 0.97) for all the developed models. The GEP produced the lowest RSME (0.01) and MAE (0.01) for L-SI. Recent studies have shown that the GEP models are effective for the prediction of water chemical quality [22, 50, 58].

In research by Ghorbani and Salehi [59], the GEP method could predicate the changes in groundwater quality data in the Isfahan Barkhaar Plain, Iran. Furthermore, the acceptable performance of the GEP model with main mathematical operators has been shown by analyzing the relationship between quality variables and river discharge [51].

Table 3. A step-by-step investigation of the effect of the water's physicochemical parameters on water corrosion and scaling indices by removing the less effective parameter of the previous step.

Index	Level	pH	TDS	T	Cl^-	SO_4^{2-}	Mg^{2+}	Ca^{2+}	CO_3^{2-}	HCO_3^-	R^2_{Test}	$R^2_{Training}$
			(mg L^{-1})	°C	(mg L^{-1})							
LSI	1	65.58	4.52	19.23				0.87	_	9.80	0.96	0.93
	2	65.47	1.35	27.60				_	_	5.58	0.91	0.96
	3	66.79	_	30.22				_	_	2.99	0.92	0.95
	4	66.19	_	33.81				_	_	_	0.94	0.94
RSI	1	49.31	6.18	38.51				_	2.09	3.68	0.88	0.90
	2	53.82	0.11	43.18				_	_	2.90	0.89	0.89
	3	52	_	44.63				_	_	3.38	0.88	0.91
	4	54.38	_	45.62				_	_	_	0.84	0.91
PSI	1		17.22	6.23				_	14.49	62.06	0.89	0.92
	2		19.01	_				_	18.56	62.34	0.89	0.91
	3		24.89	_				_	_	75.11	0.87	0.91
AI	1	55.51					11.89	3.88	_	28.71	0.96	0.94
	2	56.41					8.65	_	_	34.89	0.90	0.96
	3	60.81					_	_	_	39.19	0.89	0.94
L-SI	1				32.44	58.25			_	9.31	0.97	0.99
	2				63.32	36.68			_	_	0.97	0.97

The dash (_) indicates the removal of the low-effect parameter from the previous step.

Shah et al. [50] reported the relative superiority of the GEP compared to ANN and linear and non-linear regression models for TDS and EC of the upper Indus River basin. A MEP-based predictive model for EC and TDS water quality parameters in the upper Indus River had better generalization capabilities than traditional non-linear regression models [23].

Gholizadeh et al. [10] reported the same results for groundwater resources in the Yazd-Ardakan Plain, Iran. Therefore, the use of high-precision obtained relationships can predict the corrosion and scaling indices of water by only measuring the water's most influential physicochemical parameters. If the RMSE and MAE are closer to 0 and the correlation is closer to 1, the accuracy of GEP in predicting and modeling water corrosiveness and sedimentation indicators is higher [58]. The statistics showed that the accuracy of the models for L-SI, LSI and AI was higher than RSI and PSI. Also, RMSE, and MAE were very close to 0, their R2 values were almost equal to 1 (Table 5).

LSI is used to qualitatively assess water potential in calcium carbonate deposition formation [39]. This index shows the effect of parameters such as calcium, total alkaline, TDS, and temperature (T) in calculating saturated pH values. In other words, this index can be called the pH change required to reach water equilibrium [1]. RSI is the modified type of LSI, whose values are positive against the LSI type. This index quantitatively indicates the dependence between calcium carbonate saturation and film formation. The determinants of this index are specified by actual pH, the concentration of Ca^{2+}, HCO_3^- ions, TDS, and T [40]. LSI and RSI show the difference between the actual pH of the water and the pH saturated by calcium carbonate [41]. PSI, also known as the practical scaling index, shows the water's corrosion or scaling potential and, unlike the other three indices, is not associated with actual water pH. PSI values are influenced by water buffering capacity, and equilibrium pH is used instead of actual water pH to calculate its effect [42, 43]. AI is mainly a function of pH, calcium concentration, and alkaline [44]. The L-SI shows the effect of SO_4^{2-}, Cl^-, CO_3^{2-}, and HCO_3^- ions. CO_3^{2-} and HCO_3^- reduce corrosion, while Cl^- and SO_4^{2-} increase water corrosion [45].

In general, the present study's findings are similar to those reported by [50]), Shah et al. [50] in which they found that GEP exceeded ANN and regression models for TDS and EC. They derived two equations to illustrate the GEP model's effectiveness in monitoring river water quality. In the last few decades, the advancements in the model's accuracy have improved the prediction of water quality. Artificial intelligence (AI) approaches, including ANN and GEP, are reliable for predicting and analyzing various complex engineering issues [60].

Table 4. Linear relationships between water's corrosion and scaling indices and high-impact physicochemical parameters of water.

Index	£(x)	R^2	MSE	Max Error	P Value
LSI	1.050 pH+0.003 T-8.397	0.89	0.06	0.23	***
RSI	-1.100 pH-0.005 T+16.793	0.80	0.14	0.45	***
PSI	-0.002 TDS-0.391 HCO_3^-+8.767	0.80	0.24	0.45	***
AI	1.043 pH+0.192 HCO_3^-+3.679	0.93	0.06	0.55	***
L-SI	0.271 Cl^-+0.259 SO_4^{2-}+0.001	0.93	0.06	0.12	***

T is temperature (°C); TDS=total dissolved solids (mg L-1); HCO3-, Cl-, and SO42- the concentrations in (mg L-1)

Table 5. Obtained relationships for water's corrosion and scaling potential indices by GEP.

Index	£(x)	R^2	MAE	RMSE
LSI	$pH + \log(T)^2 - 16.096$	0.89	0.06	0.08
RSI	$\sqrt{pH}+(pH/\log T) - pH+8.485$	0.85	0.11	0.13
AI	$pH + \log(HCO_3^-) + e^{0.54} + \log(pH+HCO_3^-)$	0.90	0.04	0.07
PSI	$(4.719-(HCO_3^-)^{1/3})/\log(HCO_3^-)$	0.80	0.13	0.16
L-SI	$\sqrt{Cl^-}/6.75$	0.97	0.01	0.01

T is temperature (°C); TDS=total dissolved solids (mg L^{-1}); HCO_3^-, Cl^-, and SO_4^{2-} the concentrations in (mg L^{-1})

Regarding the management of non-linear data, ANN is a more successful method in comparison with conventional statistical techniques; therefore, prior research has indicated that ANN has a higher performance than conventional models, such as multivariate linear regression, particularly when complex interactions exist between variables [61]. Notably, ANN can accurately model complex non-linear relationships, which is crucial for precise assessments of water quality [62]. ANN is capable of effectively modeling the multiple water quality parameters' interactions. According to different approaches reviewed by [20]), ANN demonstrated reliability and acceptance in the field of river water quality modeling. Jassam et al. [63] highlighted ANN's capacity to integrate data from multiple sources, improving the accuracy of water quality index (WQI) predictions. This capability is crucial when water quality is affected in terms of numerous factors, such as temperature, salinity, and dissolved oxygen [64]. Haghiabi et al. [65] explained that ANN could be trained on historical data to provide accurate future predictions. Although regression coefficients provide useful information of the system under investigation, no procedures have been developed to extract such information from ANN parameters [20].

4. Conclusions

This study investigated the prediction of water corrosion and scaling indices using ANN and GEP models. The results showed that ANN could reasonably provide a prediction of water corrosion and scaling indices with the highest correlation coefficient (0.95, 0.91, 0.96, 0.92, and 0.99) and the lowest percentage errors (0.20, 0.44, 0.40, 0.44, and 0.08) for LSI, RSI, AI, PSI, and L-SI, respectively. The sensitivity analysis showed the strength of the pH effect as the most effective water parameter on RSI, LSI, and AI. The most effective water parameters in PSI and L-SI were HCO_3^- and SO_4^{2-}, respectively. Furthermore, GEP models introduced linear and nonlinear relationships with high precision (0.80 to 0.97). The results of this study suggested that the established GEP models with selected key parameters could be prioritized to estimate the water's corrosion and scaling potential in different climate zones of the Kermanshah Province. It is recommended that future studies introduce new GEP models for each climate zone in the study area using the new larger datasets.

Abbreviations

ANN	artificial neural networks
GEP	gene expression programming
AI	Aggressive index
LSI	Langelier saturation index
L-SI	Larson–Skold index
PSI	Puckorius scaling index
RSI	Ryznar stability index

Declaration of competing interests

The authors declare that they have no competing interests.

Authors contribution

Shabnam Vaisi: Data collection and manuscript writing, Akram Fatemi: Supervision and review, Rasool Ghobadian Supervision and review.

Funding

This research received no external funding.

Declaration of using generative AI

The authors confirm that no generative AI tools were used at any stage of the research.

References

[1] Rezaei Kalantari, R., Yari, A. R., Ahmadi, E., Azari, A., Tahmasbi Zade, M., & Gharagazlo, F. (2013). Survey of corrosion and scaling potential in drinking water resources of the villages in Qom province by use of four stability indexes (With Quantitative and qualitative analysis. Archives of Hygiene Sciences, 2(4), 127-134.

[2] Foorginezhad, S., Mohseni-Dargah, M., Firoozirad, K., Aryai, V., Razmjou, A., Abbassi, R.,...Asadnia, M. (2021). Recent advances in sensing and assessment of corrosion in sewage pipelines. Process Safety and Environmental Protection, 147, 192-213. https://doi.org/110.1016/j.psep.2020.1009.1009

[3] Davis, J. (2000). The effects and economic impact of corrosion. In Corrosion: Understanding the basics (Vol. 164, pp. 62-66). Almere, The Netherlands: ASM International.

[4] Nasr, M., & Zahran, H. F. (2014). Using of pH as a tool to predict salinity of groundwater for irrigation purpose using artificial neural network. The Egyptian Journal of Aquatic Research, 40(2), 111-115. https://doi.org/110.1016/j.ejar.2014.1006.1005.

[5] Ghaneian, M., Ehrampoush, M., Ghanizadeh, G., & Amrollahi, M. (2008). Survey of corrosion and precipitation potential in dual water distribution system in Kharanagh district of Yazd province. Tolo-e- Behdasht, 7(3-4 (25)), 65-72 (in Persian with English Abstract).

[6] Ghanizadeh, G., & Ghaneeian, M. (2009). Corrosion and precipitation potential of drinking-water distribution systems in military centers. Journal of Military Medicine, 11(3 (41)), 155-160 (in Persian with English Abstract).

[7] Świetlik, J., Raczyk-Stanisławiak, U., Piszora, P., & Nawrocki, J. (2012). Corrosion in drinking water pipes: The importance of green rusts. Water Research, 46(1), 1-10. http://dx.doi.org/10.1016/j.watres.2011.1010.1006.

[8] Aghazadeh, N., Chitsazan, M., & Golestan, Y. (2017). Hydrochemistry and quality assessment of groundwater in the Ardabil area, Iran. Applied Water Science, 7(7), 3599-3616. https://doi.org/3510.1007/s13201-13016-10498-13209.

[9] Vairavamoorthy, K., Yan, J., Galgale, H. M., & Gorantiwar, S. D. (2007). IRA-WDS: A GIS-based risk analysis tool for water distribution

systems. Environmental Modelling & Software, 22(7), 951-965. http://dx.doi.org/910.1016/j.envsoft.2006.1005.1027.

[10] Gholizadeh, A., Mokhtari, M., Naimi, N., Shiravand, B., Ehrampoush, M. H., Miri, M., & Ebrahimi, A. (2017). Assessment of corrosion and scaling potential in groundwater resources; a case study of Yazd-Ardakan Plain, Iran. Groundwater for Sustainable Development, 5, 59-65. https://doi.org/10.1016/j.gsd.2017.1004.1002.

[11] Nayeria, D., Mousavia, S. A., Darvishia, P., Mahmoudia, A., Nooria, E., & Delavaric, S. (2019). Evaluation of water quality and stability in the drinking water distribution network: A case study in the Kermanshah city, Iran. Desalination and Water Treatment, 166, 180-185. http://dx.doi.org/110.5004/dwt.2019.24623.

[12] Fatemi, A., Eskandari, Z., Gholivand, M. B., & Zarabi, M. (2022). Corrosion and scaling potentials of rural water distribution network in different climate zones of Kermanshah province, Iran. Journal of Applied Research in Water and Wastewater, 9(2), 1-9. https://doi.org/10.22126/arww.22022.25678.21187.

[13] Siddha, S., & Sahu, P. (2022). Evaluation of corrosivity and scaling properties of groundwater of Central Gujarat for industrial usage. Arabian Journal of Geosciences, 15(9), 891. https://doi.org/810.1007/s12517-12022-10197-12510.

[14] Alborzi, M., & Abtahi, S. A. R. (2006). Corrosion rate prediction with artificial neural network case: Crude oil distillation overhead systems. Industrial Management Studies, 4(13), 41-66. (in Persian).

[15] Afan, H. A., Yafouz, A., Birima, A. H., Ahmed, A. N., Kisi, O., Chaplot, B., & El-Shafie, A. (2022). Linear and stratified sampling-based deep learning models for improving the river streamflow forecasting to mitigate flooding disaster. Natural Hazards, 112(2), 1527-1545. https://doi.org/1510.1007/s11069-11022-05237-11067.

[16] Chaplot, B., & Birbal, P. (2022). Development of stage-discharge rating curve using ANN. International Journal of Hydrology Science and Technology, 14(1), 75-95. https://doi.org/10.1504/IJHST.2022.123643.

[17] Alaie, M., Zanguie, M., & Zanguie, H. (2015). Forecasting the water quality of Neyshabur plain using MLP networks and adaptive neural fuzzy inference system (ANFIS). National Conference of Environmental Sciences and Engineering, Ahvaz, Iran, 17 February.

[18] Mohammadi, A. A., Ghaderpoori, M., Yousefi, M., Rahmatipoor, M., & Javan, S. (2016). Prediction and modeling of fluoride concentrations in groundwater resources using an artificial neural network: a case study in Khaf. Environmental Health Engineering and Management, 3(4), 217-224. http://dx.doi.org/210.15171/EHEM.12016.15123

[19] Emami, S., Noruzi-Sarkarabad, R., & Choopan, Y. (2021). Use of artificial neural network and imperialist competitive algorithm to evaluate the groundwater quality of Jolfa Plain for various uses. Amirkabir Journal of Civil Engineering, 53(1), 313-330. https://doi.org/310.22060/ceej.22018.14258.25605.

[20] Rajaee, T., Khani, S., & Ravansalar, M. (2020). Artificial intelligence-based single and hybrid models for prediction of water quality in rivers: A review. Chemometrics and Intelligent Laboratory Systems, 200, 103978. http://dx.doi.org/103910.101016/j.chemolab.102020.103978.

[21] Aldrees, A., Khan, M. A., Tariq, M. A. U. R., Mustafa Mohamed, A., Ng, A. W. M., & Bakheit Taha, A. T. (2022). Multi-expression programming (MEP): water quality assessment using water quality indices. Water, 14(6), 947. http://dx.doi.org/910.3390/w14060947.

[22] Kulisz, M., & Kujawska, J. (2021). Application of artificial neural network (ANN) for water quality index (WQI) prediction for the river Warta, Poland. Journal of Physics: Conference Series, 2130. IOP Publishing Ltd., https://dx.doi.org/10.1088/1742-6596/2130/1081/012028.

[23] Alqahtani, A., Shah, M. I., Aldrees, A., & Javed, M. F. (2022). Comparative assessment of individual and ensemble machine learning models for efficient analysis of river water quality. Sustainability, 14(3), 1183. https://doi.org/1110.3390/su14031183.

[24] Beiranvand, S., & Sahraiean, K. (2019). Types of neural networks and their applications The Forth International Conference on Management and Humanities Researches, Tehran, Iran, 14 March.

[25] Angabini, S., Ahmadi, H., Feiznia, S., Vaziri, B., & Ershadi, S. (2014). Suspended sediment concentration estimation using artificial neural networks and fuzzy rule base model case study: Jagin dam. Journal of Applied Sciences Research, 10(14), 12-17.

[26] Muzzammil, M., Alama, J., & Danish, M. (2015). Scour prediction at bridge piers in cohesive bed using gene expression programming. Aquatic Procedia, 4, 789-796. https://doi.org/710.1016/j.aqpro.2015.1002.1098.

[27] Singley, J., Beaudet, B., & Markey, P. (1984). Corrosion manual for internal corrosion of water distribution systems.

[28] Rossum, J. R., & Merrill, D. T. (1983). An evaluation of the calcium carbonate saturation indexes. Journal American Water Works Association, 75(2), 95-100. https://doi.org/110.1002/j.1551-8833.1983.tb05075.x.

[29] Dastorani, M. T., Azimi Fashi, K., Talebi, A., & Ekhtesasi, M. R. (2013). Estimation of suspended sediment using artificial neural network (case study: Jamishanwatershed in Kermanshah). Journal of Watershed Management Research, 3(6), 61-74.

[30] Menhaj, M. (2021). Computational Intelligence-Volume I: Fundamentals of Neural Networks. Amirkabir University of Technology Publications, 716.

[31] Reddy, S. (2003). Estimation of watershed runoff using artificial neural networks PhD thesis, IARI, Division of Agricultural Engineering, New Delhi].

[32] Vesta Services, I. (2000). Qnet2000 Shareware. Vesta Services, Inc.

[33] Bagatur, T., & Onen, F. (2014). A predictive model on air entrainment by plunging water jets using GEP and ANN. KSCE Journal of Civil Engineering, 18(1), 304-314. http://dx.doi.org/310.1007/s12205-12013-10210-12207.

[34] Ferreira, C. (2005). Gene Expression Programming and the Evolution of Computer Programs. In L. Nunes de Castro & F. J. Von Zuben (Eds.), Recent Developments in Biologically Inspired Computing (pp. 82-103). IGI Global. https://doi.org/10.4018/978-1-59140-312-8.ch005

[35] Wilson, S. (2008). Classifier conditions using gene expression programming (No. IlliGAL Report No. 2008001): University of Illinois at Urbana-Champaign. Proc. 32nd Australasian Computer Science Conference (ACSC 2009), JWellington, New Zealandanuary, 19-23.

[36] Ferreira, C. (2006). Gene expression programming: mathematical modeling by an artificial intelligence (Vol. 21). Springer.

[37] Lopes, H. S., & Weinert, W. R. (2004). EGIPSYS: an enhanced gene expression programming approach for symbolic regression problems. International Journal of Applied Mathematics and Computer Science, 14(3), 375-384.

[38] Gepsoft, L. (2009). Discover the excellent of modelling with GeneXpro Tools 4.0, the best and most intutive modeling software in the market. (Accessed 27 Oct 2009) http://www.gepsoft.com/.

[39] Antony, A., Low, J. H., Gray, S., Childress, A. E., Le-Clech, P., & Leslie, G. (2011). Scale formation and control in high pressure membrane water treatment systems: A review. Journal of Membrane Science, 383(1-2), 1-16. http://dx.doi.org/10.1016/j.memsci.2011.1008.1054.

[40] Rafferty, K. (1999). Scaling in geothermal heat pump systems. US Department of Energy Idaho Operations Office785 DOE Place Idaho Falls, ID 83401.

[41] Amiri, M. C. (2018). Water treatment principles (2nd ed.). Arkan Danesh Publication (in Persian).

[42] Davil, M. F., Mahvi, A. H., Norouzi, M., Mazloomi, S., Amarluie, A., Tardast, A., & Karamitabar, Y. (2009). Survey of corrosion and scaling potential produced water from Ilam water treatment plant. World Applied Science Journal, 7(11), 11-24.

[43] Pourzamani, H., Ghazavi, M., & Samani, A. (2005). Evaluation of drinking water quality considering corrosion in industrial park of Oshtorjan, Isfahan 8th National Congress on

Environmental Health, Tehran, Iran, 8 November.

[44] Piri Elm, R., Shams, G., Shahmansuri, M., & Farzadkia, M. (2008). Survey of corrosion and scaling potential in drinking water of distribution system of Khoramshahr City with corrosion Index. Quartery Newsletter, 10(13), 15 (in Persian with English Abstract).

[45] Larson, T. E., & Skold, R. V. (1958). Laboratory studies relating mineral quality of water to corrosion of steel and cast iron. Corrosion, 14(6), 43-46.

[46] Zai, K. M., Nawaz, R., Bhatti, I. A., & Bhatti, H. N. (2008). A pHs equation for calcium carbonate scale prediction in cooling water systems. Journal of the Chemical Society of Pakistan, 30(2), 182-185.

[47] Ebrahimi, A., Kamarehie, B., Asgari, G., Seid, M. A., & Roshanaei, G. (2012). Drinking water corrosivity and sediment in the distribution network of Kuhdasht, Iran. Health System Research, 8(3), 480-486 (in Persian with English Abstract).

[48] Taghavi, M., Mohammadi, M. H., Radfard, M., Fakhri, Y., & Javan, S. (2019). Assessment of scaling and corrosion potential of drinking water resources of Iranshahr. MethodsX, 6, 278-283. https://doi.org/10.1016/j.mex.2019.02.002

[49] Amouei, A., Asgharnia, H., Fallah, H., Yari, A. R., & Mahmoudi, M. (2017). Corrosion and Scaling Potential in Drinking Water Distribution of Babol, Northern Iran Based on the Scaling and Corrosion Indices. Archives of Hygiene Sciences, 6, 1-9.

[50] Shah, M. I., Javed, M. F., & Abunama, T. (2021). Proposed formulation of surface water quality and modelling using gene expression, machine learning, and regression techniques. Environmental Science and Pollution Research, 28, 13202-13220. http://dx.doi.org/13210.11007/s11356-13020-11490-13209.

[51] Dehghani, R., & Mahmoudi, S. (2014). Performance assessment of gene expression programming model related to chemical parameters to modeling river discharge (case study: Simineh Rood) The First International Congress of Earth Sciences Tehran, Iran, 18 February.

[52] Moasheri, S. (2013). Estimation of the spatial distribution of groundwater quality of Birjand Plain with geostatistical–artificial neural network combined method University of Zabol].

[53] Abbasi Maedeh, P., Mehrdadi, N., Bidhendi, G. N., & Abyaneh, H. Z. (2013). Application of artificial neural network to predict total dissolved solids variations in groundwater of Tehran Plain, Iran. International Journal of Environmental Sustainability, 2(1), 10-20.

[54] Gholabi, M., & Karami, B. (2013). Simulation and prediction of water qualitative parameters using artificial neural network, fuzzy-neural method, and statistical regression (case study: Karoun River, Khouzestan Province) The Ninth International Conference on River Engineering, Ahvaz, Iran, 22 January.

[55] Mirsanjari, M. M., Mohammadyari, F., Basiri, R., & Hamidi pour, F. (2017). Modeling quality parameters EC, SAR and TDS in groundwater using artificial neural network (case study: Mehran Plain and Dehloran). Human and Environment, 15(3), 1-12 (in Persian with English Abstract).

[56] Mirzavand, M., Ghasemieh, H., Akbari, M., & Sadatinejad, S. J. (2015). Artificial neural network (ANN) model for ground water quality simulation (Case study: Kashan aquifer). Journal of Range and Watershed Management, 68(1), 159-171. https://doi.org/110.22059/jrwm.22015.53892.

[57] Chen, K., Chen, H., Zhou, C., Huang, Y., Qi, X., Shen, R.,...Ren, H. (2020). Comparative analysis of surface water quality prediction performance and identification of key water parameters using different machine learning models based on big data. Water Research, 171, 115454. http://dx.doi.org/115410.111016/j.watres.112019.115454.

[58] Birbal, P., Azamathulla, H., Leon, L., Kumar, V., & Hosein, J. (2021). Predictive modelling of the stage–discharge relationship using Gene-Expression Programming. Water Supply, 21(7), 3503-3514.

http://dx.doi.org/3510.2166/ws.2021.3111.

[59] Ghorbani, M., & Salehi, A. (2011). The use of gene expression programming to examine changes in groundwater quality data with water level fluctuations in the Barkhor Plain of Isfahan 6th National Congress on Civil Engineering, Semnan, Iran, 25 April.

[60] Shah, M. I., Alaloul, W. S., Alqahtani, A., Aldrees, A., Musarat, M. A., & Javed, M. F. (2021). Predictive modeling approach for surface water quality: Development and comparison of machine learning models. Sustainability, 13(14), 7515. http://dx.doi.org/7510.3390/su13147515.

[61] Zare Abyaneh, H. (2014). Evaluation of multivariate linear regression and artificial neural networks in prediction of water quality parameters. Journal of Environmental Health Science and Engineering, 12(1), 40. https://doi.org/10.1186/2052-1336X-1112-1140.

[62] Yan, J., Liu, J., Yu, Y., & Xu, H. (2021). Water quality prediction in the Luan River based on 1-DRCNN and BiGRU hybrid neural network model. Water, 13(9), 1273. https://doi.org/1210.3390/w13091273.

[63] Jassam, J. D., Mawlood, I. A., & Sayl, K. N. (2024). Applying remote sensing and artificial neural networks for water quality index modeling in the Euphrates River. International Journal of Design & Nature and Ecodynamics, 19(2), 605-611. https://doi.org/610.18280/ijdne.190226.

[64] Vasanthi, S. S., & Kumar, S. A. (2019). Application of artificial neural network techniques for predicting the water quality index in the Parakai Lake, Tamil Nadu, India. Applied Ecology & Environmental Research, 17(2), 1947-1958. http://dx.doi.org/1910.15666/aeer/11702_19471958.

[65] Haghiabi, A. H., Nasrolahi, A. H., & Parsaie, A. (2018). Water quality prediction using machine learning methods. Water Quality Research Journal, 53(1), 3-13. https://doi.org/10.2166/wqrj.2018.2025.

How to cite this paper:

Fatemi, A., Vaisi, S. & Ghobadian, R. (2025). Water's corrosion and scaling potential prediction using artificial neural networks and gene expression programming in several rural water distribution networks in Kermanshah Province, Iran. Advances in Environmental Technology, 11(4), 396-413. DOI: 10.22104/aet.2025.7194.1980