Environmental Earth Sciences (2023) 82:395
https://doi.org/10.1007/s12665-023-11059-y

ORIGINAL ARTICLE

Conjunct application of machine learning and game theory in groundwater quality mapping

Ali Nasiri Khiavi[1] · Mohammad Tavoosi[1] · Alban Kuriqi[2]

Received: 18 March 2023 / Accepted: 15 July 2023 / Published online: 9 August 2023

Abstract

Groundwater quality (GWQ) monitoring is one of the best environmental objectives due to recent droughts and urban and rural development. Therefore, this study aimed to map GWQ in the central plateau of Iran by validating machine learning algorithms (MLAs) using game theory (GT). On this basis, chemical parameters related to water quality, including K^+, Na^+, Mg^{2+}, Ca^{2+}, SO_4^{2-}, Cl^-, HCO_3^-, pH, TDS, and EC, were interpolated at 39 sampling sites. Then, the random forest (RF), support vector machine (SVM), Naive Bayes, and K-nearest neighbors (KNN) algorithms were used in the Python programming language, and the map was plotted concerning GWQ. Borda scoring was used to validate the MLAs, and 39 sample points were prioritized. Based on the results, among the ML algorithms, the RF algorithm with error statistics MAE = 0.261, MSE = 0.111, RMSE = 0.333, and AUC = 0.930 was selected as the most optimal algorithm. Based on the GWQ map created with the RF algorithm, 42.71% of the studied area was in poor condition. The proportion of this region in the classes with moderate and high GWQ was 18.93% and 38.36%, respectively. The results related to the prioritization of sampling sites with the GT algorithm showed a great similarity between the results of this algorithm and the RF model. In addition, the analysis of the chemical condition of critical and non-critical points based on the results of RF and GT showed that the chemical aspects, carbonate balance, and salinity at critical points were in poor condition. In general, it can be said that the simultaneous use of MLA and GT provides a good basis for constructing the GWQ map in the central plateau of Iran.

Keywords Artificial intelligence · Borda scoring algorithm · Hydrogeochemistry · Ion balance diagram · Optimal decision making

Introduction

Water resources management (WRM) is one of the most important challenges at the global level (Abu El-Magd et al. 2023). Water is considered the most important human need for social, economical, and agricultural development (Siebert et al. 2010; Singh et al. 2012; Kubicz et al. 2021). On the other hand, the quantity and quality of water and drinking water consumption sustain lakes and wetlands and directly affect biodiversity (Khan et al. 2023). Groundwater resources generally provide about 50% of drinking water and 40% of industrial needs globally (Udmale et al. 2014). Today, groundwater resources face many problems threatening their quantity and quality (Burri et al. 2019; El Asri et al. 2019; Houéménou et al. 2020). Therefore, the different characteristics of groundwater resources are affected and lead to the destruction of these valuable reserves, which have become a major crisis in different regions (Eissa et al. 2016; Eid et al. 2023).

In Iran, as in other developing countries, groundwater quality (GWQ) is seriously threatened by excessive exploitation of groundwater resources, extensive use of chemicals, and pesticide intrusion (Panaskar et al. 2016; Akhtar et al. 2021). On the other hand, industrialization and population growth have accelerated groundwater pollution (Kumar et al. 2019a; Sarker et al. 2021). Thus, industrial and domestic

✉ Alban Kuriqi
alban.kuriqi@tecnico.ulisboa.pt

Ali Nasiri Khiavi
ali.nasirikhiavi@modares.ac.ir

Mohammad Tavoosi
m_tavosi@modares.ac.ir

1 Department of Watershed Management Engineering, Faculty of Natural Resources and Marine Sciences, Tarbiat Modares University, Tehran, Iran

2 CERIS, Instituto Superior Técnico, Universidade de Lisboa, Av. RoviscoPais 1, 1049-001 Lisbon, Portugal

wastewater has entered the water table and caused water pollution (Ojekunle et al. 2020). Determining parameters for the optimal use of groundwater resources are groundwater's chemical composition and biology (Davraz and Oezdemir 2014; Mallick et al. 2018). Groundwater resources change due to climatic and lithological conditions and processes caused by human activities (Abbasnia et al. 2019; Jehan et al. 2019). The presence of fluoride sulfate, nitrate, and the presence of metals such as calcium, magnesium, sodium, manganese, cadmium, nickel, chromium, and arsenic in water above the permissible limit causes problems for various uses, including drinking water supply and agriculture (Rashid et al. 2020; Tyagi and Sarma 2020).

Traditional methods of GWQ assessment are often expensive and time-consuming. Meanwhile, implementing machine learning algorithms (MLAs) can help adopt long-term strategies regarding the GWQ and how to use it by predicting and evaluating water quality indicators (Masoud et al. 2022; Noori et al. 2022). In some studies, several approaches such as machine learning algorithms (MLAs) and water quality index (WQI) have been used to evaluate water quality (Mohamed et al. 2014; Ramadan 2016; Adimalla et al. 2018; Hamlat and Guidoum 2018; Ahmed et al. 2020; Badeenezhad et al. 2020; Suvarna et al. 2020; El-Magd et al. 2021, 2022).

Among the parameters that help to evaluate the GWQ are total dissolved solids (TDS), Mg^{2+}, Na^{+}, HCO_3^{-}, and sodium absorption ratio (SAR) (Abu El-Magd et al. 2023). Some studies have evaluated the effect of human manipulation on groundwater pollution (Safa et al. 2020; Ravish et al. 2021). New and more popular models and techniques have been proposed for GWQ assessment. These methods are data-oriented and have higher accuracy. These methods have been used in several studies, such as support vector machines (SVM) and artificial neural networks (ANN) (Pei-Yue et al. 2011; El-Magd et al. 2020). Some studies based on MLAs, such as adaptive neuro-fuzzy inference system (ANFIS) and SVM, have evaluated the GWQ (Elsayed et al. 2021; Tao et al. 2022; Eid et al. 2023; Sahour et al. 2023).

The investigation of the research background showed that the GWQ had been investigated using different tools and models. In addition, WQI and sometimes MLAs have been widely used to investigate groundwater resources and GWQ. Although in many studies, only the modeling of GWQ has been addressed, and the validation of the methods and models used has not been widely investigated. Therefore, the study sought to complete the current research deficiency and validated MLAs using the GT algorithm. Game theory is one of the most important optimal methods of multi-criteria decision-making, which has low uncertainty in choosing the best criteria and alternatives compared to other multi-criteria decision-making methods. In addition, this study investigated and comprehensively analyzed the chemical status and important ions related to critical points in terms of GWQ that MLA and GT identified.

One of the main reasons for choosing the province of Chaharmahal and Bakhtiari was the heavy use of groundwater in the form of wells and springs in this region. Meanwhile, it was necessary to study the groundwater resources in this region, which account for a significant proportion of water resources and require optimal and integrated management. Therefore, this study aimed to map GWQ in the central plateau of Iran by validating MLAs, including random forest (RF), SVM, Naive Bayes, and K-nearest neighbors (KNN) algorithms using Borda scoring algorithm based on GT.

Materials and methods

Description of the study area

Chaharmahal and Bakhtiari province, with an area of 18,122 km^2, is a highland covering the central plateau of Iran. Chaharmahal and Bakhtiari province is a region where almost 80% of the city is covered by mountains and hills (Rahimi 2012; Zamani-Ahmadmahmoodi et al. 2019). These mountains have 16 peaks with an altitude of more than 3500 m. The morphology of the province includes a regular alternation of northwestern and southeastern elevations separated by plains with the same trend. The average slope of the province is 42%, and more than 58% of the area has a slope of 30% or more. Precipitation in the region is often influenced by Mediterranean atmospheric currents and low Sudanese atmospheric pressure (Lashkari et al. 2021).

The average annual rainfall of the province is 560 mm (Arab Amiri and Mesgari 2019). Chaharmahal and Bakhtiari provinces have about 10% of the country's water resources. The country’s two major and strategic rivers, the Karon and the Zainderud, originate here. One of the main problems of this province is that a high percentage of groundwater is used for agriculture. In addition, a large percentage of drinking water is obtained from aquifers (Heshmati and Beigi 2012). The geographical location and sampling points are shown in Fig. 1.

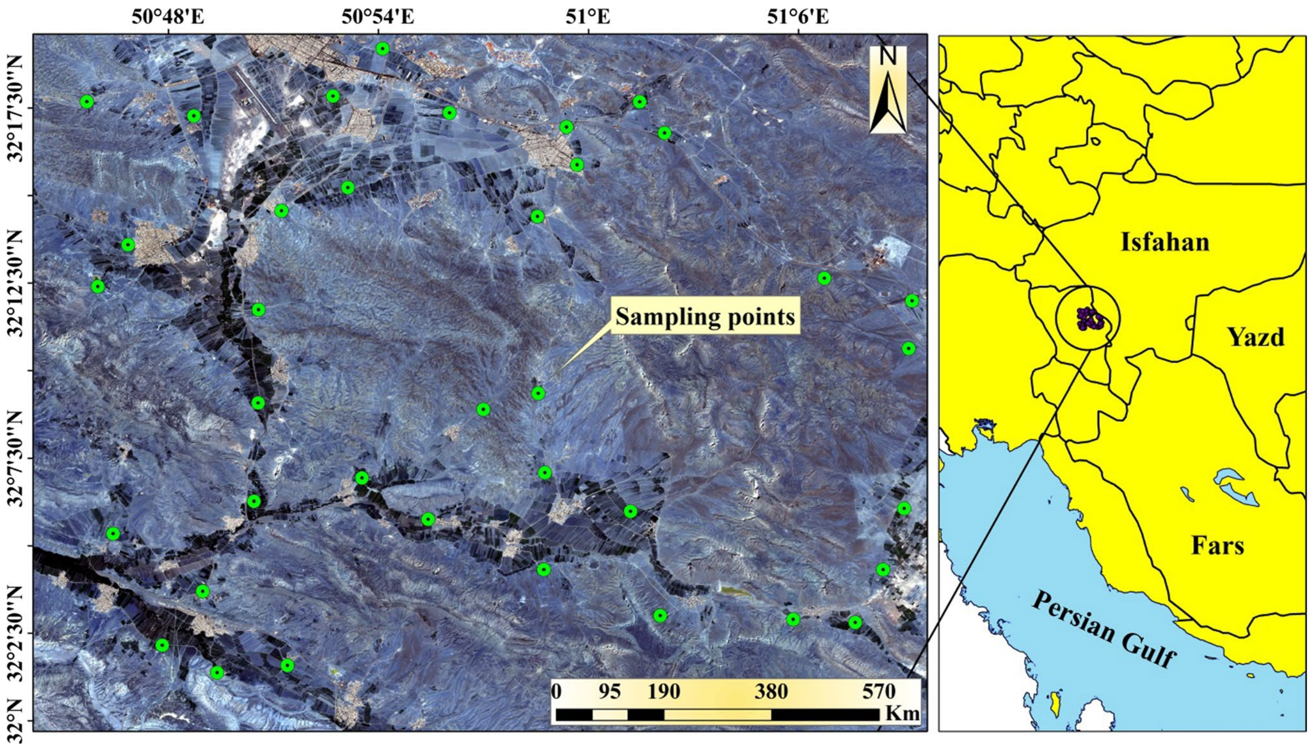

Fig. 1 A view of the studied area in the central plateau of Iran (Sentinel-2; 2022)

Table 1 Some characteristics of sampling points are Chaharmahal and Bakhtiari Province, Iran

Sampling points	X UTM	Y UTM	Elevation (m)
1	514,535	3,562,592	2212
2	517,037	3,553,338	2113
3	490,785	3,575,910	2067
4	489,226	3,568,569	2040
5	479,360	3,565,561	2046
6	514,195	3,551,645	2139
7	478,661	3,550,326	2012
8	512,004	3,545,626	2129
9	486,246	3,567,346	2043
10	499,023	3,571,760	2120
11	480,855	3,544,428	2028
12	509,209	3,545,787	2120
13	478,000	3,563,362	2085
14	498,003	3,548,407	2060
15	513,253	3,548,415	2142
16	492,813	3,551,065	2020
17	482,678	3,547,274	2016
18	482,324	3,572,362	2059
19	501,899	3,551,483	2087
20	486,477	3,543,356	2056
21	497,755	3,557,717	2164
22	495,290	3,556,864	2128
23	484,985	3,552,032	1998
24	514,381	3,560,073	2205
25	497,727	3,567,040	2145
26	477,526	3,573,127	2182
27	499,499	3,569,768	2152
28	510,588	3,563,785	2375
29	498,056	3,553,523	2067
30	488,564	3,573,404	2046
31	489,841	3,553,251	2021
32	493,782	3,572,511	2059
33	483,319	3,542,985	2025
34	485,185	3,557,220	2014
35	515,983	3,547,591	2136
36	503,233	3,545,981	2219
37	502,309	3,573,096	2158
38	485,205	3,562,129	2060
39	503,421	3,571,452	2177

Data sources and analyses

GWQ data are for 39 sampling sites in Chaharmahal and Bakhtiari, Iran. Water quality parameters included K^+, Na^+, Mg^{2+}, Ca^{2+}, $SO4^{2-}$, Cl^-, HCO_3^-, pH, TDS, and EC (Asghari et al. 2018; Iqbal et al. 2018). These data were provided by Iran Water Resources Management Company (IWRMC). Some characteristics of the sampling sites are shown in Table 1. Table 2 also contains the values of quality factors at each point. Sentinel 2 satellite imagery, digital elevation models (DEM), and shapefiles for the sampling points were used to monitor the studied area.

Table 2 Quantitative values of groundwater quality conditioning factors, Chaharmahal and Bakhtiari Province, Iran

Sampling points	K[a]	Na[a]	Mg[a]	Ca[a]	SO_4[a]	Cl[a]	HCO_3[a]	pH	TDS[a]	EC
1	0.02	0.64	1.70	2.35	1.02	0.51	3.11	8.11	300.07	462.64
2	0.02	1.31	1.57	2.17	1.47	1.00	2.49	8.15	331.53	510.21
3	0.02	0.91	1.06	2.29	0.75	0.41	3.07	7.91	276.55	428.64
4	0.03	1.11	1.86	2.52	0.86	1.01	3.61	7.83	347.35	534.47
5	0.02	0.18	1.04	2.57	0.32	0.29	3.14	7.87	238.07	366.55
6	0.03	0.63	1.74	2.45	0.75	0.81	3.24	8.11	311.12	478.86
7	0.03	0.37	2.03	2.82	0.37	0.36	4.52	7.84	327.93	505.03
8	0.03	1.62	2.54	3.50	1.07	1.95	4.65	7.81	494.14	759.77
9	0.03	1.53	3.02	3.46	2.16	2.16	3.66	7.90	510.53	787.50
10	0.02	0.59	1.35	2.85	0.86	0.47	3.45	7.88	303.21	466.93
11	0.02	0.17	0.82	2.68	0.22	0.25	3.21	8.00	234.40	360.42
12	0.02	0.38	0.88	2.64	0.34	0.48	3.07	7.85	247.96	381.65
13	0.02	0.10	0.76	2.46	0.21	0.20	2.86	7.85	207.38	319.14
14	0.02	0.20	1.11	2.69	0.25	0.29	3.44	7.90	250.90	386.40
15	0.03	1.95	1.93	3.38	0.59	3.77	2.92	8.01	477.25	734.17
16	0.03	0.95	2.17	3.22	0.82	1.15	4.35	7.84	404.39	622.17
17	0.04	0.26	1.78	2.52	0.35	0.39	3.85	7.91	287.13	441.77
18	0.03	1.01	1.90	2.96	1.06	0.96	3.82	8.03	371.94	588.19
19	0.02	0.26	1.74	2.54	0.36	0.32	3.80	8.08	288.32	444.21
20	0.02	0.39	1.74	2.06	0.54	0.31	3.27	8.09	265.75	408.57
21	0.03	0.69	2.08	3.72	1.00	0.70	4.81	7.90	417.62	637.69
22	0.03	1.05	1.50	2.49	0.65	1.49	2.90	7.90	318.48	488.06
23	0.03	0.47	2.38	3.15	0.73	0.83	4.46	7.88	376.43	579.29
24	0.03	0.81	1.50	2.48	0.94	0.77	3.03	8.07	315.04	479.00
25	0.03	1.53	1.99	3.02	1.30	1.89	3.32	7.86	416.00	640.18
26	0.03	0.55	1.93	2.52	0.59	1.02	3.39	7.91	317.54	489.08
27	0.02	0.40	1.07	2.93	0.58	0.39	3.44	7.85	278.42	428.56
28	0.02	0.23	0.97	1.92	0.38	0.32	2.38	8.03	196.94	303.01
29	0.03	0.77	1.95	2.98	0.88	1.22	3.58	7.85	363.81	559.83
30	0.02	0.66	1.82	3.14	0.78	0.52	4.34	7.86	353.81	544.52
31	0.03	0.68	2.35	2.90	0.81	0.75	4.32	7.83	372.99	572.55
32	0.02	0.79	1.40	2.70	0.89	0.51	3.47	7.90	311.44	478.97
33	0.03	0.34	1.84	3.24	0.57	0.45	4.38	7.88	342.50	526.27
34	0.02	0.27	1.43	2.75	0.21	0.31	3.90	7.83	279.00	427.89
35	0.03	2.17	3.27	4.26	1.17	5.33	3.17	8.03	646.25	990.07
36	0.02	0.15	0.85	2.67	0.22	0.25	3.17	7.84	229.09	351.31
37	0.02	0.53	1.32	2.84	0.70	0.45	3.51	7.86	293.82	452.97
38	0.04	3.02	3.07	4.17	1.55	4.65	4.07	7.96	659.09	1014.11
39	0.02	0.50	1.58	2.51	0.86	0.38	3.31	7.97	291.59	448.56

[a]K, Na, Mg, Ca, SO_4, Cl, HCO_3, TDS (Mg L^{-1}); EC (μS cm^{-1})

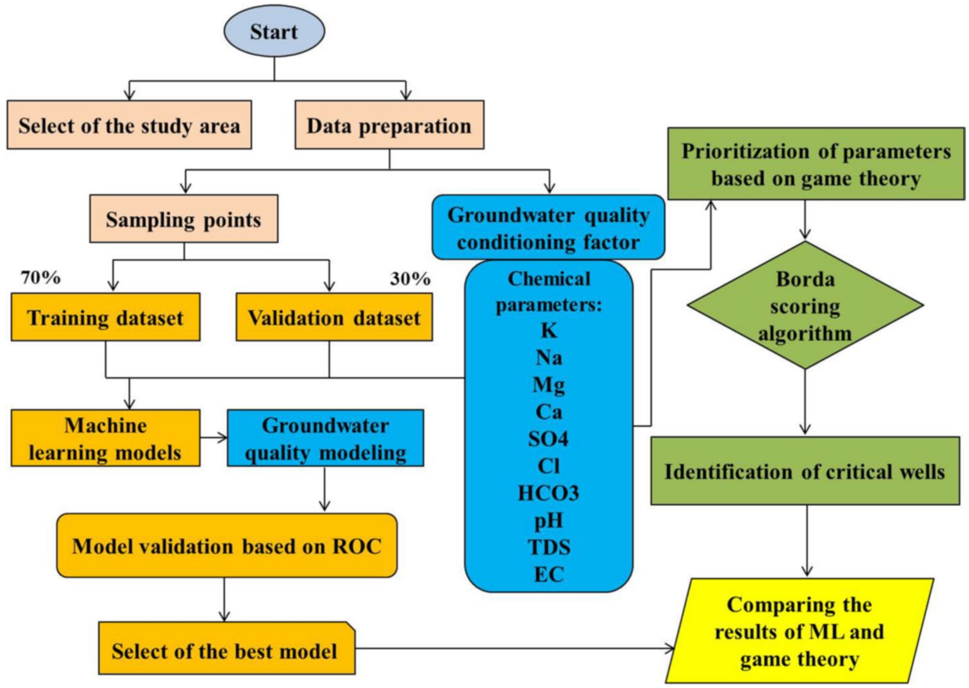

Fig. 2 Flowchart of research methodology

Research methodology

Based on the flow chart (Fig. 2), groundwater quality parameters were quantified at each sampling point to conduct this study. Then, in ArcGIS 10.8 software, the values of each parameter were interpolated using the Kriging method, and all ten parameters were converted into grids (Ali and Ahmad 2020). These maps related to GWQ conditioning factors were shown in Fig. 3. Then, MLAs including RF, SVM, Naive Bayes, and KNN were used to construct the GWQ map (Leong et al. 2021; Ramadhani et al. 2021; Wang et al. 2021; Ilić et al. 2022). These models were coded and implemented based on Python (Lee et al. 2020). In this way, 70% of the data set was used for training and 30% for validation. Finally, the most optimal model was selected based on the error statistics, including mean absolute error (MAE), MSE (mean square error), root mean square error (RMSE), and receiver operating characteristics (ROC) (Koranga et al. 2022; Rasool et al. 2022).

To compare MLA results, the GT algorithm was used (Padarian et al. 2020). According to the Borda scoring algorithm (Khiavi et al. 2022), ten studied parameters (K^+, Na^+, Mg^{2+}, Ca^{2+}, $SO_4{}^{2-}$, Cl^-, $HCO_3{}^-$, pH, TDS, and EC) were prioritized at 39 sampling sites. The most critical points were selected in terms of GWQ. Finally, after selecting the high, medium, and low-quality points based on GWQ, AqQA software (Wu et al. 2017) was used, and the status of the points with the highest and lowest critical conditions in terms of water type, density, CO_2 content, carbonate balance, and irrigation water was examined. Explanations of the application of MLA, GT, and the software used are provided below:

Machine learning algorithms

MLA algorithms have been used to predict continuous numerical outputs (Fernández-Delgado et al. 2019). MLA is divided into three categories: supervised MLA, unsupervised MLA, and reinforcing MLA (Vafakhah et al. 2022). This study used RF, SVM, Naive Bayes, and KNN.

RF algorithm This method is a non-parametric algorithm (Noi et al. 2017; Zhou et al. 2019). It is one of the most optimal ML methods for decision-making and in a supervised manner (Breiman 2001; Vorpahl et al. 2012). This method consists of three parts: the number of parameters used in each tree, the number of trees, and the number of nodes (Peters et al. 2008).

In this model, the random vector θ_k, which is independent of the random vectors $\theta_1 \ldots .\theta_{k-1}$, is generated for the K tree. Also, all vectors have the same distribution. The regression tree grows using training data and θ_k. The set of k trees is equal to $\{h_1(x), h_2(x) \ldots h_k(x)\}$, which is $\{h_k(x) = h(x, \theta_k), x = \{x_1, x_2, \ldots x_p\}\}$ here. These vectors are the next P input vector that makes up a forest. Generated K outputs correspond to each tree equal to $\hat{y}_1 = h_1(x), \hat{y}_2 = h_2(x), \ldots \hat{y}_{k=}h_k(x)$, where $\hat{y}_k$ is the output of the K tree. The average of all tree predictions is calculated to obtain the final output.

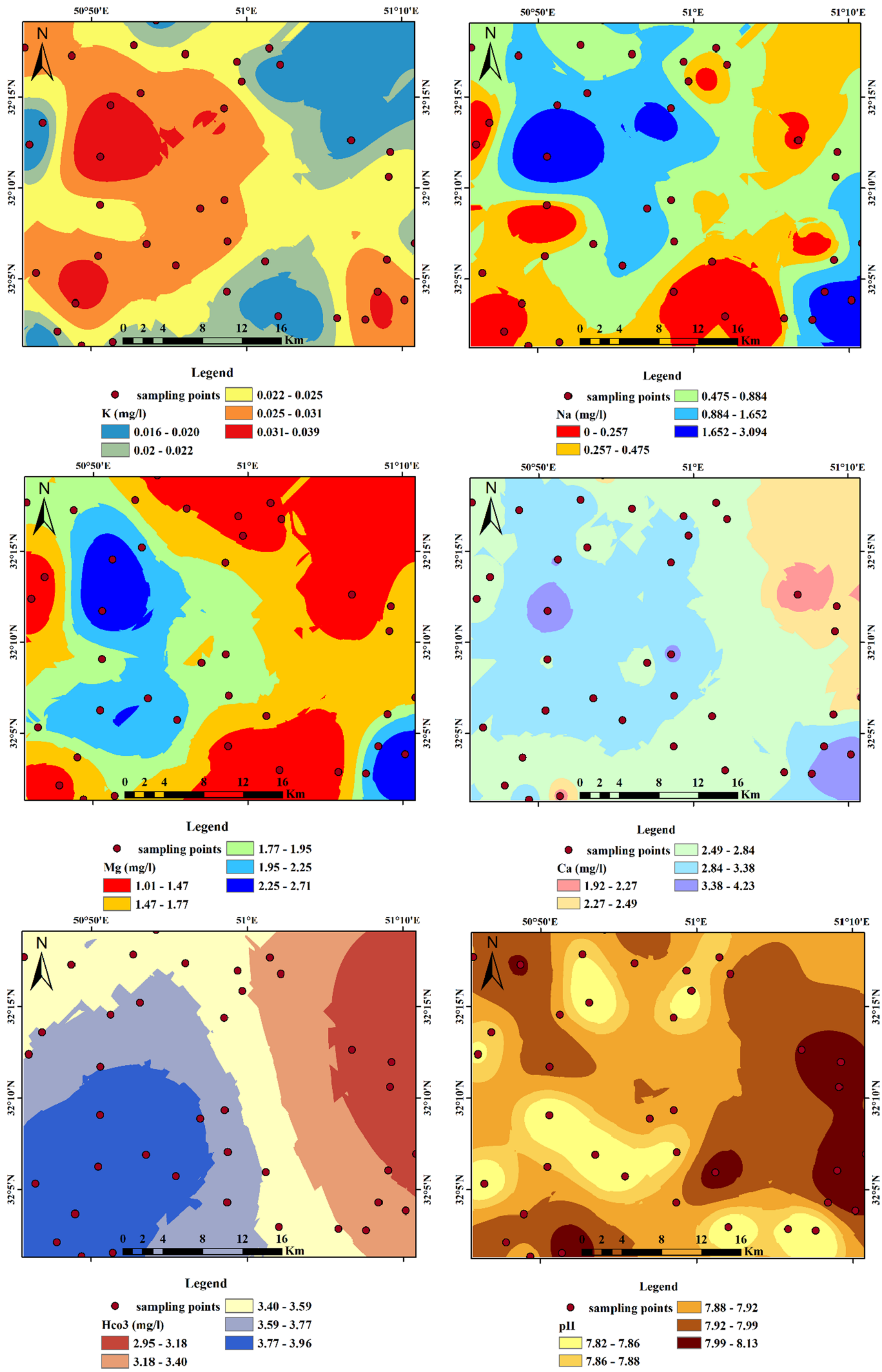

Fig. 3 Groundwater quality conditioning factors, Chaharmahal and Bakhtiari Province, Iran

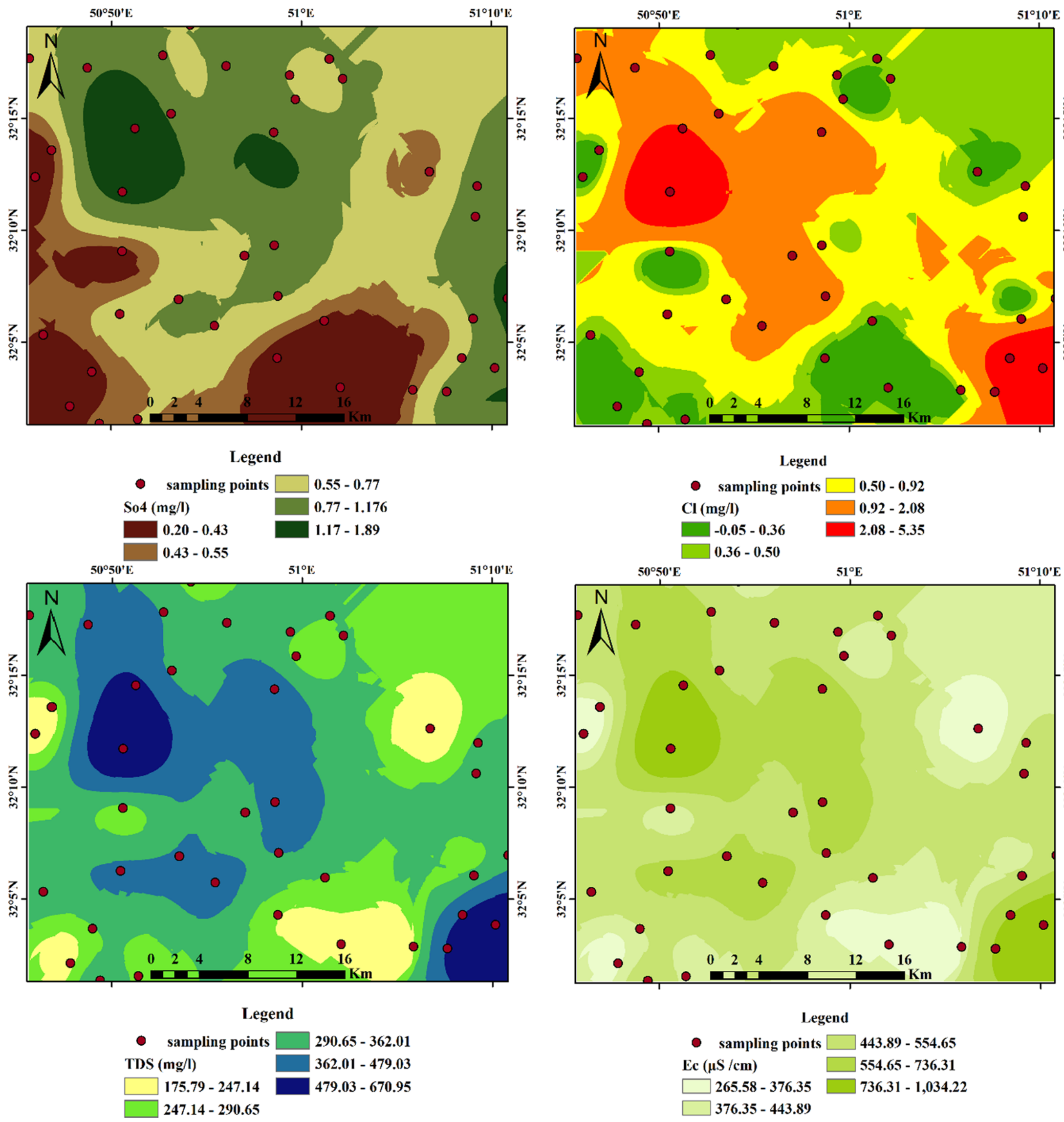

Fig. 3 (continued)

SVM algorithm This algorithm was developed 1995 as a decision model (Cortes and Vapnik 1995). It belongs to the supervised learning methods used for classification and regression. The basis of SVM is a linear classification of data. In this algorithm, we try to choose the line with the most reliable margin (BOTSIS et al. 2011). In machine learning, support vector machines with associated learning methods have been supervised to analyze the data for optimal decisions (Sharifi Garmdareh et al. 2018; Vafakhah and Khosrobeigi Bozchaloei 2020).

Suppose N samples of the population are given by $X \in R^m$, $\left\{X_K, Y_K\right\}_{K=1}^{N}$, $Y \in R$, then Eq. (1) can be a regression function as below:

$$f(x) = W\varnothing(X) + b, \tag{1}$$

where $\varnothing$ represents the kernel functions, X is an input factor with m elements, Y is the output parameter, W is a weight vector, and b represents an error (Shabani et al. 2016). Cortes and Vapnik (1995) showed optimization of the following

Eq. (2) where ε_k and ε_k^* are parameters to reduce training bias based on Eq. (3) (Xie et al. 2012; Shabani et al. 2016):

$$\frac{1}{2}\|W\|^2 + C\sum_{k=1}^{k=N}(\varepsilon_k - \varepsilon_k^*) \tag{2}$$

$$\text{subjectto}\begin{Bmatrix} Y_k - W^T\varnothing\left(X_k\right) - b \le \epsilon + \varepsilon_k \\ W^T\varnothing(X_k + b - Y_k \le \epsilon + \varepsilon_k^*) \\ \varepsilon_k\varepsilon_k^* \ge 0 \end{Bmatrix} k = 1, 2, \ldots, N \tag{3}$$

Naive Bayes This algorithm is a supervised method that uses Bayes' theorem. This means that the method only assumes that each input variable is independent. This model works well despite insufficient or inappropriate data (Osisanwo et al. 2017). Depending on the data type of each characteristic, a different method is required. More specifically, the data estimate parameters from one of three standard probability distributions (Frank et al. 2000; Ren et al. 2009). A polynomial distribution can be used for categorical variables such as numbers or labels. If the variables are binary, the binomial distribution can be used. The Gaussian distribution is often used for a numeric variable, such as a measurement. This classification algorithm requires less data volume and is more efficient than others (Tsangaratos and Ilia 2016).

KNN algorithm This algorithm is used for the classification of unknown problems assuming the presence of certain features (*X*) and a certain value (*Y*) (Avand et al. 2019). The KNN is a non-parametric model. To calculate the actual value to the nearest points, the maximum number of *K* in each neighbor and each selected point is used (Betrie et al. 2013). This algorithm assumes that the neighboring cells are classified into the same class based on the density function and the target matrix.

The basis of this model is to calculate the proximity of the real-time prediction value $X\_r = \{x\_1n, x\_2n, x\_3n, \ldots, x\_mr\}$ to the forecast value in each observation $X\_t = \{x\_1b, x\_2b, x\_3b, \ldots, x\_mr\}$ and based on the Euclidean distance function (D_{rt}) (Naghibi et al. 2020):

$$D_{rt} = \sqrt{\sum_{i=1}^{m} w_i(x_{ir} - x_{it})^2}, t = 1, 2, \ldots, n, \tag{4}$$

where w_i ($i = 1, 2, \ldots, m$) is the weight of predictors whose sum is equal to one.

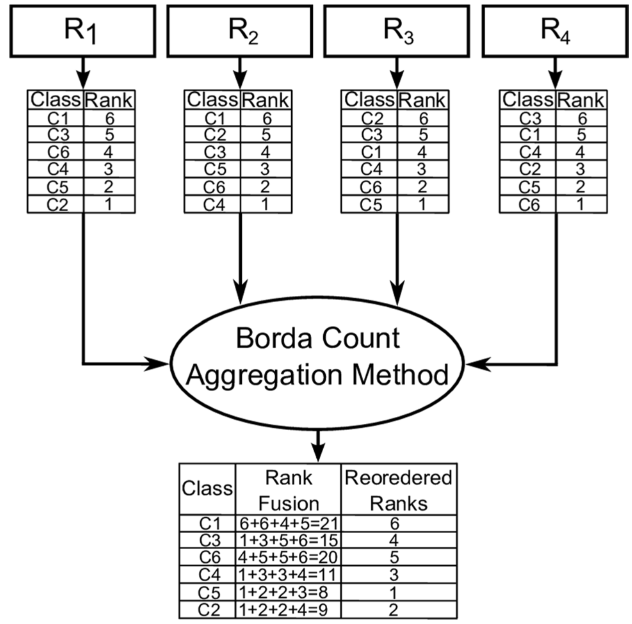

Fig. 4 Simple schematic of the GT algorithm (Source: Avand et al. 2021)

Borda scoring algorithm

After quantifying the GWQ parameters, this algorithm was applied to weigh the parameters and identify areas at risk. For this purpose, the quantitative values of each parameter at each sampling site were calculated. Then, the output matrix of the GT algorithm was created (Fig. 4). Finally, the sampling sites were classified into three categories high, medium, and low quality based on the game theoretic algorithm. The Borda scoring algorithm was a prioritization method for group decisions to evaluate parameters. The Borda score was determined for each candidate. It was the sum of the individual scores for each parameter (Avand et al. 2021). For every n representative, there are n ranks. Accordingly, $n - 1$ points were assigned to the representative with the first rank and $n - 2$ with the second rank, and so on (Khiavi et al. 2022). The weight of each candidate was denoted by *BS* (*A*) and noted as (Eqs. 5–8):

$$BS(A) = (n-1) * \#\{i|i\text{ranks}A\text{first}\} + (n-2) \tag{5}$$

$$* \#\{i|i\text{ranks}A\text{second}\} + \cdots + 1 \tag{6}$$

$$* \#\{i|i\text{ranks}A\text{secondtolast}\} + 0 \tag{7}$$

$$* \#\{i|i\mathrm{ranks}A\mathrm{last}\}, \quad (8)$$

where *#{…}* is the number of parameters (Balinski and Laraki 2007; Adhami and Sadeghi 2016).

Application of AqQA software

Water chemists have written AqQA for water chemists. The AqQA software can perform six data homogeneity tests based on AWWA-1030-E standard methods. The AqQa tools can easily create 11 diagrams, including time series, Schoeller diagrams, ion balance, Durov, Piper, and Stiff. The main advantages of this software were mixing the sample during the simulation process, determining the equilibrium of anions and cations using the ion balance diagram, determining the chemical properties of water, determining the type of water, and calculating the main properties of liquids (dissolved solids measurement, density measurement, and others) (Ebrahime Moghadam and Abbasnejad 2020). In addition, carbon balance calculations, TDS, EC, and organic, inorganic, biological, isotopic, and radioactive analyzes make the AqQA platform a suitable tool for water quality data analysis (Salehi et al. 2016; Gholamrazai 2020).

To summarize the research methodology, a GWQ map was first created based on chemical parameters and using MLA in Python. Then, using different error statistics, the best model was selected. Then, Borda scoring was applied to identify critical points related to GWQ. Finally, AqQA software analyzed water quality conditions at critical and non-critical points. GWQ conditions were analyzed chemically and qualitatively.

Results

Results of machine learning algorithms (MLAs)

The results related to the error statistics for determining the best MLA are shown in Table 3. In addiiton, the receiver operating characteristics for evaluating the GWQ maps at each sampling point based on machine learning are shown in Fig. 5.

The significant percentage of GWQ conditioning factors (K^+, Na^+, Mg^{2+}, Ca^{2+}, SO_4^{2-}, Cl^-, HCO_3^-, pH, TDS, and EC) based on the fit of the best model (RF algorithm) is shown in Fig. 6. Figure 7 also shows the GWQ map based on the RF algorithm (with the lowest error and highest AUC).

Table 3 Predictive capability of groundwater quality modeling using MLA

Models	Criteria			
	MAE	MSE	RMSE	AUC
RF	0.261	0.111	0.333	0.930
SVM	0.375	0.370	0.612	0.620
Naive Bayes	0.25	0.25	0.5	0.770
KNN	0.458	0.450	0.677	0.610

Results of the GT algorithm

After creating the decision matrix for an optimal MCDM, the Borda scoring algorithm based on GT was used, and the GWQ conditioning factors were prioritized based on sample points. In this algorithm, the criteria, GWQ conditioning factors, and alternatives were sample points, and the prioritization results are included in Table 4. To compare the prioritization and identification of critical sampling points based on GWQ in three classes, Table 5 was used. The prioritization results were presented using the Borda scoring algorithm based on GT and the most optimal MLA (RF algorithm) (Table 5).

Results of the chemical properties of the fluid

The results of the analysis of chemical parameters (K^+, Na^+, Mg^{2+}, Ca^{2+}, SO_4^{2-}, Cl^-, HCO_3^-, pH, TDS, and EC) affecting GWQ were determined using AqQA software at critical and non-critical sampling points (the highest and lowest scores based on the RF algorithm and the Game theoric Borda scoring algorithm) and presented in Table 6. The results of the ratio of important ions, including Na^+/Cl^-, $Ca^{2+}+Mg^{2+}/HCO_3^-+SO_4^-$, Ca2+/$HCO_3^-$, Mg^{2+}/HCO_3^-, Ca^{2+}/SO_4^- and SO_4^-/HCO_3^-, were shown in Fig. 8. The ion balance diagram for critical and non-critical points (the highest and lowest scores based on RF algorithm and Game theoric Borda scoring algorithm) in the study area, Iran, is shown in Fig. 9.

Discussion

The change in GWQ, usually caused by mismanagement of water harvesting, chemical fertilizers, and similar factors, has become a prelude to destroying other resources, directly or indirectly (Li et al. 2018; Kumar et al. 2019b). In many countries, including Iran, groundwater is one of the most important water sources for drinking, industry, and agriculture. Utilizing these sources and surface water containment has always been considered an option. In many parts of Iran, groundwater is essential for drinking, agricultural, and industrial water supply due to a lack of access to surface

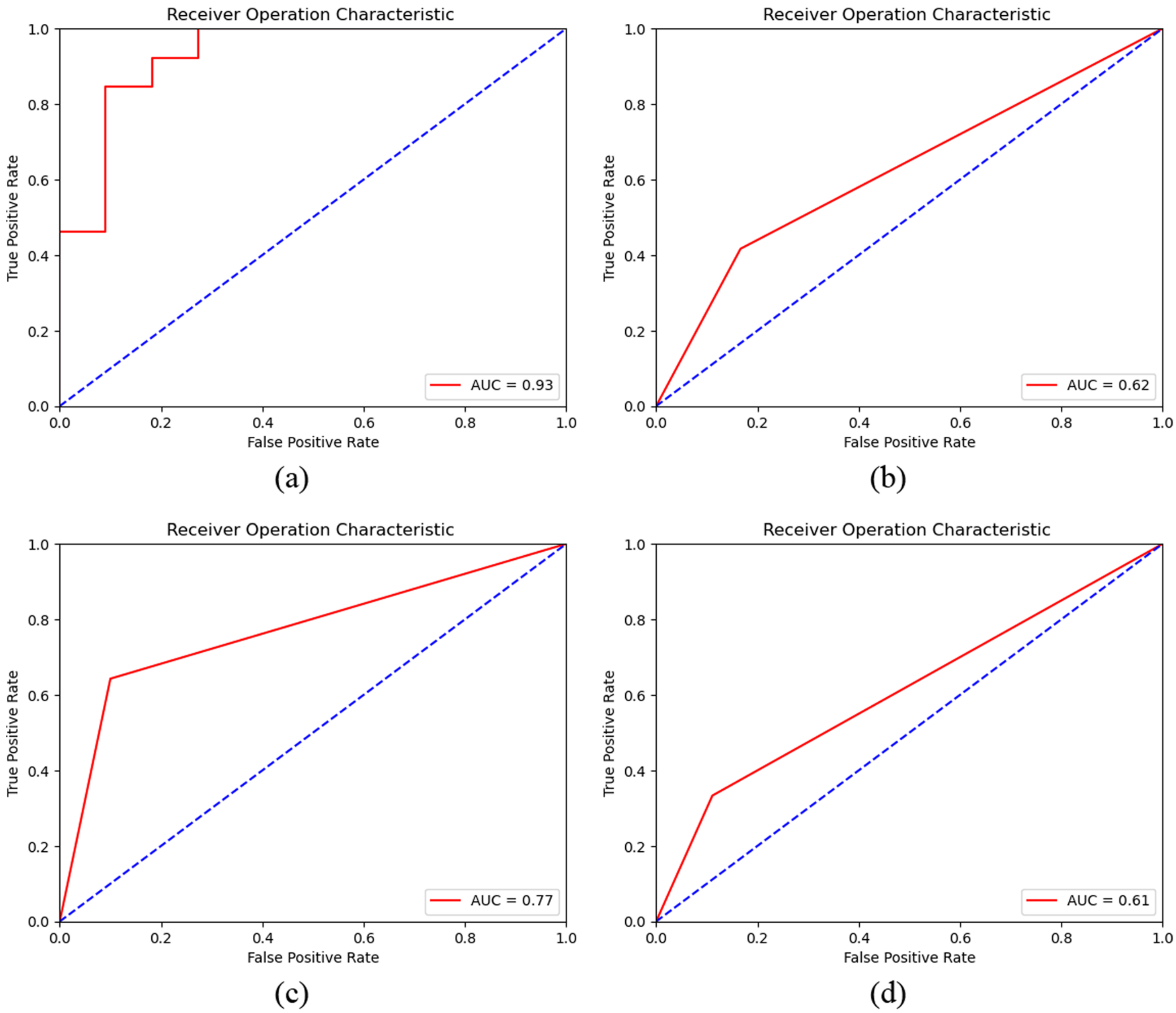

Fig. 5 ROC curve for groundwater quality mapping: **a** RF, **b** SVM, **c** Naive Bayes, **d** KNN

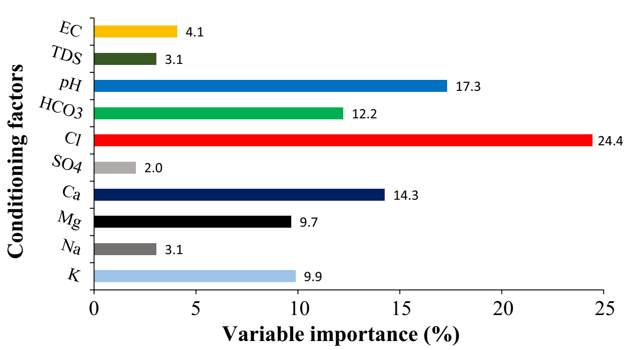

Fig. 6 Importance of groundwater quality conditioning factors using the RF model

water (Mirzaei et al. 2019; Maghrebi et al. 2020). Applying appropriate management methods in using existing water resources and reducing the high cost of their development and use can also optimize the scale of use of these resources. In this study, MLA and GT were combined to investigate GWQ in the central plateau of Iran.

According to Table 3 and Fig. 5, among the MLA, including RF, SVM, Naive Bayes, and KNN, the RF algorithm with error values of MAE = 0.261, MSE = 0.111, RMSE = 0.333, and AUC = 0.930 was selected as the most optimal algorithm in GWQ mapping (Tesoriero et al. 2017; Norouzi and Moghaddam 2020; Nafouanti et al. 2021; He et al. 2022). In addition, the RF algorithm confirmed several fields (Lianjun

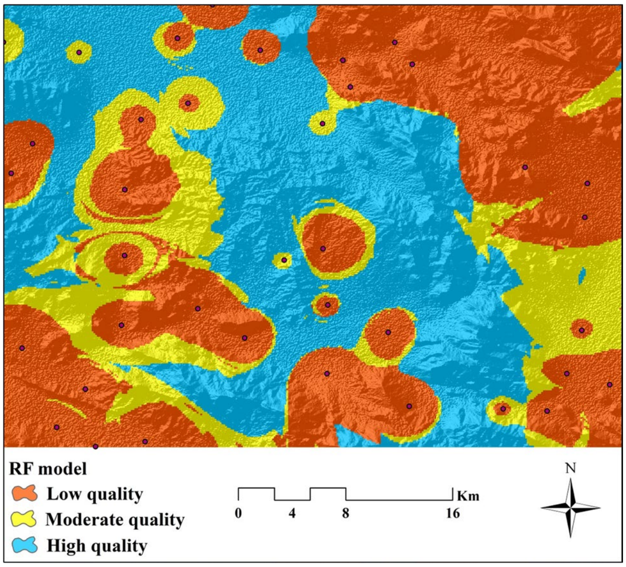

Fig. 7 Groundwater quality mapping based on the RF model, Chaharmahal and Bakhtiari Province, Iran

2016; Pham et al. 2021; Vafakhah et al. 2022). In addition, many studies highlighted the use of MLA for WRM (Starzyk 2010), landslides (Hong et al. 2015), erosion (Tien Bui et al. 2019), and urban water management (Rozos 2019). The main advantage of this algorithm over other MLAs is that this algorithm uses the best variables randomly selected from the input variables to build a tree (Demir and Sahin 2022; Vafakhah et al. 2022). This procedure reduces the overall error of the model. Another advantage of this algorithm is that it avoids model fitting. This algorithm's insensitivity to the data's normality is another important advantage. On the other hand, it provides suitable results for classified data types and generally has a reasonable application speed compared to other MLAs (Kotsiantis and Pintelas 2004; Rahman et al. 2020).

Among the factors affecting GWQ and based on the importance of these factors, chlorine ions (Rao et al. 2012; Krishna Kumar et al. 2015) with 25% and sulfate ions (Subramani et al. 2005; Sharma and Kumar 2020) with 2.5% had the greatest and least influence on GWQ in the central plateau of Iran, respectively (Fig. 6). Based on the GWQ map using the algorithm RF (Fig. 7), 42.71% (558.12 ha) of the studied area in the central plateau of Iran was in poor condition. The percentage of classes with moderate and high GWQ in this region was 18.93% (247.35 ha) and 38.36% (501.33 ha), respectively (Ahankoub et al. 2022). In general, the results of the RF algorithm showed that parts of the central plateau of Iran were in a critical condition in terms of GWQ (Bhunia et al. 2018; Jamshidzadeh and Barzi 2018; Esfandiari et al. 2019; Talebiniya et al. 2019). In this context, (Mousavi et al. 2020) investigated the spatial and temporal changes of GWQ parameters based on drinking water and agriculture in the Lordegan Plain of Chaharmahal and Bakhtiari provinces, Iran. The results showed that TH and the TDS parameters were more beneficial for drinking water consumption. In addition, for agriculture, SAR and EC parameters were very good in the whole plain during the statistical period.

After mapping GWQ based on chemical parameters using MLAs, the GT algorithm was used to validate the RF algorithm to measure the accuracy of ML in identifying critical areas (Table 4). Based on the results of GT, sampling point 30, with a score of 356, was selected as the most critical point in terms of GWQ. In addition, point number 38, with a score of 20, was selected as the most non-critical area in the central plateau of Iran. Based on the ML results, these two sampling points gave similar results to the algorithm GT (Table 5). The results of GT showed that using different parameters was very important for checking GWQ and proved the necessity of MCDM methods (Srdjevic et al. 2012). According to Madani (2010), game-theoretic algorithms are one of the best methods to evaluate the decision-making of stakeholders and policymakers in a region. The Borda scoring algorithm considers majority opinion to determine the degree of importance (Elkind et al. 2011; Mahjouri and Bizhani-Manzar 2013). This algorithm was easy to use, which increased its popularity. Adhami and Sadeghi (2016) and Mahjouri and Bizhani-Manzar, (2013) found this method suitable for studies in which the priorities of the majority of voters are considered. Of course, GT had many advantages, but one of the main problems of this method was its semi-distribution, which was not pixel-oriented, unlike MLA (Avand et al. 2021).

After determining the most critical and non-critical points using various methods, the chemical conditions of these areas were examined using GWQ. The results presented in Table 6 confirm the results of RF and Borda scoring based on GT. Thus, the water quality condition at sampling site 30, the most critical area, was unfavorable. Thus, at this site, Mg-Cl (Zakaria et al. 2021) and the TH and TDS (Sarath Prasanth et al. 2012; Tiwari and Singh 2014; Aryafar et al. 2019; Karthik et al. 2019) were about 53% higher than at point number 38 (the most uncritical site, based on the results of RF and Borda scoring algorithm). On the other hand, carbonate balance also exhibited high variability (Morgenstern and Daughney 2012; Singh et al. 2013). In addition, the evaluation of GWQ about irrigation showed

Table 4 Weighted scores of sampling points based on groundwater quality conditioning factors, Borda algorithm

Sampling points	Groundwater quality conditioning factors									
	K^+	Na^+	Mg^{2+}	Ca^{2+}	SO_4^-	Cl^-	HCO_3^-	pH	TDS	EC	Total
1	25	10	30	21	9	9	36	5	22	22	189
2	12	7	19	13	8	8	26	34	12	12	151
3	36	33	36	34	38	35	25	20	36	36	329
4	8	12	20	1	11	6	14	35	7	7	121
5	32	14	34	30	18	25	35	16	30	29	263
6	31	30	13	8	16	32	3	19	21	20	193
7	11	4	8	18	4	3	18	21	6	6	99
8	27	24	28	27	27	31	23	10	27	27	251
9	24	23	31	36	30	21	38	18	33	32	286
10	35	35	35	35	33	34	37	0	35	35	314
11	10	11	3	15	6	17	7	8	5	5	87
12	22	22	33	24	22	22	31	3	29	28	236
13	34	36	26	33	14	36	4	29	34	34	280
14	21	26	14	7	29	23	5	33	19	19	196
15	30	37	38	38	34	38	11	32	37	37	332
16	13	20	17	4	31	19	8	37	15	15	179
17	7	32	15	2	36	27	1	38	23	23	204
18	23	19	18	5	20	24	13	36	17	17	192
19	5	5	4	0	10	7	0	30	0	0	61
20	6	1	2	16	2	1	10	4	2	2	46
21	4	2	1	17	3	2	12	28	3	3	75
22	37	6	21	12	7	12	28	25	11	11	170
23	19	9	23	32	12	14	34	15	24	24	206
24	16	8	12	20	0	5	29	2	10	8	110
25	18	25	11	19	28	18	21	22	18	18	198
26	3	27	6	3	19	13	6	23	8	10	118
27	29	17	27	10	15	29	17	24	20	21	209
28	28	31	24	11	26	28	24	1	25	25	223
29	17	21	22	29	21	20	32	11	26	26	225
30	38	38	37	37	37	37	30	26	38	38	356
31	1	15	16	9	24	10	15	27	13	13	143
32	15	16	9	22	17	15	22	13	14	14	157
33	33	34	29	28	35	33	16	12	32	33	285
34	9	13	7	25	13	11	19	7	9	9	122
35	14	18	10	23	25	16	20	17	16	16	175
36	20	29	25	26	32	26	27	31	28	30	274
37	26	28	32	31	23	30	33	6	31	31	271
38	0	0	0	6	1	0	2	9	1	1	20
39	2	3	5	14	5	4	9	14	4	4	64

that the salinity risk at point number 30 was much higher than in other areas, and the sodium absorption ratio (SAR) was also in an inappropriate condition (Rawat et al. 2018) (Table 6). In this context, (Gharechaee et al. 2022) evaluated the vulnerability of groundwater to salinity in the southern plains of the Bakhtegan watershed in central Iran and concluded that a large part of the studied area was vulnerable to high salinity. The ion ratio between sampling sites also showed a high correlation between the values of sodium and chlorine ions, about 0.83. In China, (Zhang et al. 2021) also

Table 5 Comparative results of prioritizing sampling points based on GTA and RF

Sampling points	Borda algorithm	RF model
1	Low quality	Low quality
2	Low quality	Low quality
3	Low quality	Low quality
4	Moderate	Low quality
5	Low quality	Low quality
6	Low quality	Moderate
7	Moderate	Low quality
8	Low quality	Low quality
9	Low quality	Low quality
10	Low quality	Low quality
11	Moderate	Low quality
12	Low quality	Low quality
13	Low quality	Low quality
14	Low quality	Low quality
15	Low quality	Low quality
16	Low quality	Low quality
17	Low quality	Low quality
18	Low quality	Low quality
19	Moderate	Low quality
20	High quality	Low quality
21	Moderate	Low quality
22	Low quality	Lowquality
23	Low quality	Low quality
24	Moderate	Low quality
25	Low quality	Low quality
26	Moderate	Moderate
27	Low quality	Low quality
28	Low quality	Low quality
29	Low quality	Low quality
30	Low quality	Low quality
31	Moderate	Low quality
32	Low quality	Low quality
33	Low quality	Moderate
34	Moderate	Low quality
35	Low quality	Low quality
36	Low quality	Moderate
37	Low quality	Low quality
38	High quality	High quality
39	Moderate	Low quality

concluded that there was a high correlation between Na and Cl ions to study GWQ and ion ratios. This was although no significant trend was observed for the other ratios (Fig. 8). the ion balance diagram also showed that the number of ions affecting water quality was very high at the critical sampling point compared to other areas.

Conclusion

Increasing pollution due to population growth, urban wastewater discharge, industrial and agricultural wastewater disposal, and landfills have contributed to the spread of pollution and the degradation of water resources. Therefore, this study aimed to map the GWQ in the central plateau of Iran by validating the MLAs using GT algorithms. On this basis, chemical parameters related to water quality, including K^+, Na^+, Mg^{2+}, Ca^{2+}, SO_4^{2-}, Cl^-, HCO_3^-, pH, TDS, and EC, were interpolated at 39 sampling sites. Then the algorithms RF, SVM, Naive Bayes, and KNN in Python were used. The map in terms of GWQ was presented in three classes (high, moderate, and low quality). The Borda scoring algorithm was used to validate the MLA, and 39 sample points were prioritized. Finally, AqQA software was used, and the critical and non-critical points were analyzed based on the results of MLA and GT according to chemical aspects, carbonate balance, ionic ratios, and salinity. Based on the results, the RF algorithm was selected as the most optimal algorithm for GWQ mapping among the MLA algorithms. The results of the RF algorithm showed that parts of the central plateau of Iran are in a critical condition concerning GWQ. The results of GT showed that using different parameters was very important for checking GWQ and proved the necessity of MCDM methods.

The results related to the prioritization of sampling sites using the GT algorithm showed a high similarity between the results of this algorithm and the RF model in GWQ mapping. In addition, the analysis of the chemical status of the critical and non-critical points in terms of water quality based on the results of RF and GT showed that the chemical aspects, carbonate balance, SAR, HCO3- content, and salinity hazard at the critical points (based on two methods of MLA and GT) were in poor condition. In addition, the ion ratio between sampling points showed a high correlation between the values of sodium and chlorine ions. In general, it can be said that the combined application of MLA and GT based on the results of this study provides a good basis for the construction of the GWQ map in the central plateau of Iran. Due to human intervention, the development of unauthorized wells, and change in climatic components, groundwater quantity, and quality have decreased in Chaharmahal and Bakhtiari provinces of Iran. The results of this study can also help policymakers in managing groundwater resources. For future studies, it is suggested to use new deep learning algorithms and optimal MCDM methods, such as the best–worst method (BWM). With more complete and comprehensive data, GWQ should be studied in other parts of the central plateau of Iran.

Table 6 Analysis of different parameters of GWQ in critical and non-critical points based on Borda scoring algorithm based on GT

Data analysis	WQP	Number of sampling points	
		Sample 30	Sample 38
Fluid properties	Water type	Mg-Cl	Ca-HCO_3
	TDS	529 mg/l	208 mg/kg
	Density	0.99743 g/cm^3	0.99719 g/cm3
	EC	815 μmho/cm	320 μmho/cm
	TH (as $CaCO_3$)	18.846 mg/l	9.5369 mg/kg
Carbonate equilibrium	CO_3	237.1×10^{-6} mmolal	173.1 × 10–6 mmolal
	HCO_3	0.04001	0.04758
	CO_2	748.8×10^{-6}	0.001424
	Partial Pressure of CO_2	20.41×10^{-6} atm	38.81×10^{-6} atm
Irrigation waters	Salinity hazard	High	Medium
	Sodium Adsorption Ratio	230×10^{-3}	8.44×10^{-3}

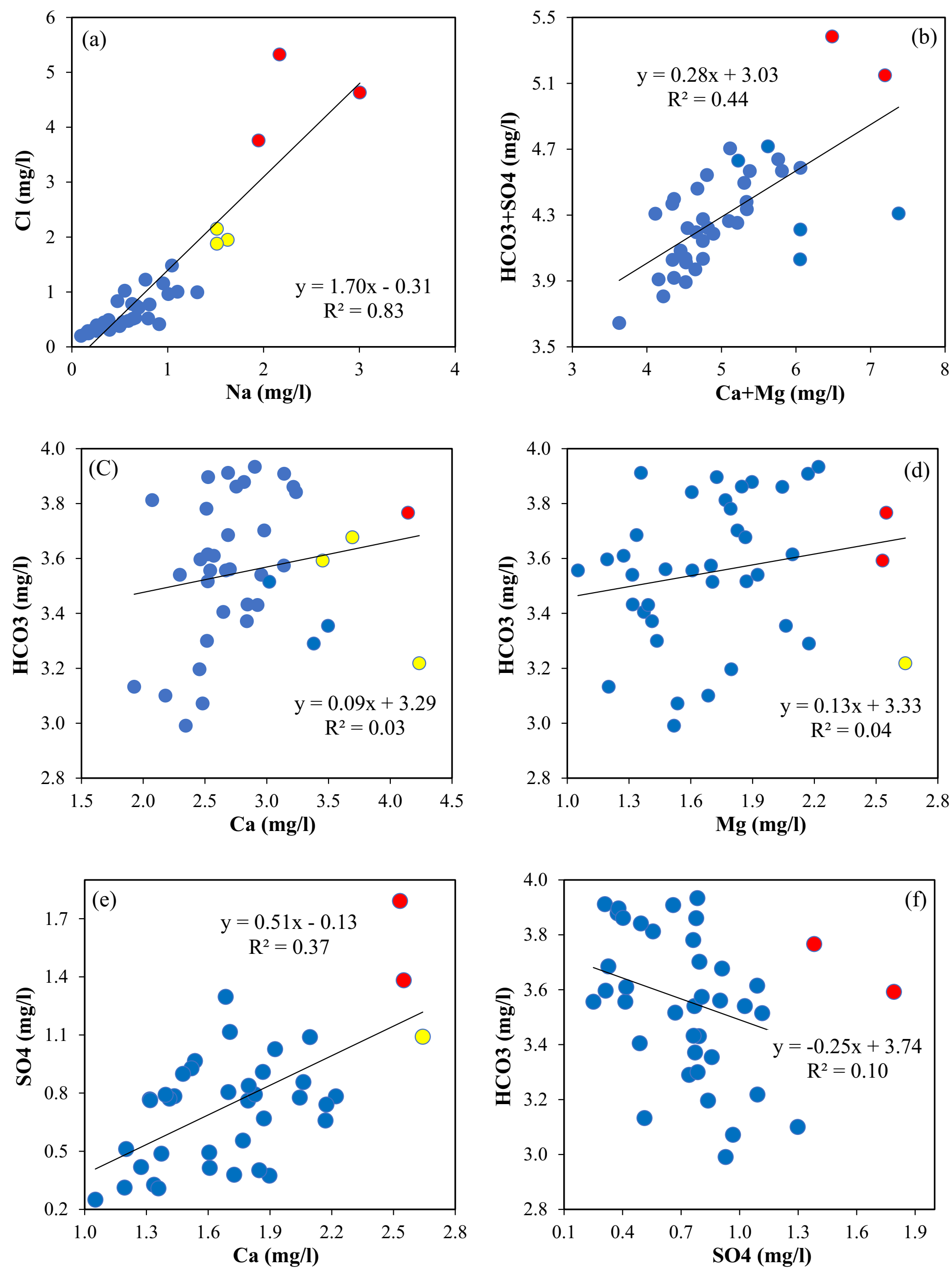

Fig. 8 Ionic ratio plots of the major ions: **a** Na/Cl, **b** Ca + Mg/HCO_3 + SO_4, **c** Ca/HCO_3, **d** Mg/HCO_3, **e** Ca/SO_4, **f** SO_4/HCO_3

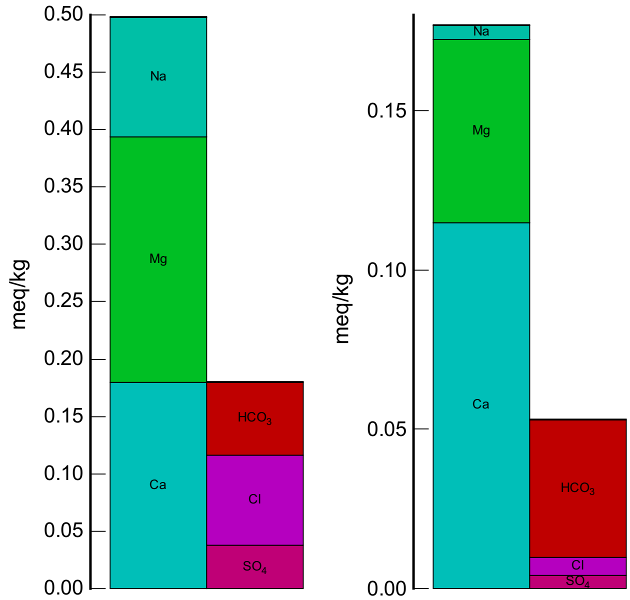

Fig. 9 Ion balance diagram related to critical (left) and non-critical (right) points

Author contributions Conceptualization, ANK, and MT; methodology, ANK, and MT; validation, ANK, MT, and AK; formal analysis, ANK, and MT, investigation, ANK, MT, and AK; writing—original draft preparation, ANK and MT; writing—review and editing, ANK, MT, and AK; supervision, ANK, and AK. All authors have read and agreed to the published version of the manuscript.

Funding Open access funding provided by FCT|FCCN (b-on).

Data availability We have no permission to release data and codes.

Declarations

Conflict of interest The authors declare no conflict of interest.

References

Abbasnia A, Ghoochani M, Yousefi N et al (2019) Prediction of human exposure and health risk assessment to trihalomethanes in indoor swimming pools and risk reduction strategy. Hum Ecol Risk Assess an Int J 25:2098–2115

Abu El-Magd SA, Ismael IS, El-Sabri MAS et al. (2023) Integrated machine learning-based model and WQI for groundwater quality assessment: ML, geospatial, and hydro-index approaches. Environ Sci Pollut Res 30(18):1–14

Adhami M, Sadeghi SH (2016) Sub-watershed prioritization based on sediment yield using game theory. J Hydrol 541:977–987. https://doi.org/10.1016/j.jhydrol.2016.08.008

Adimalla N, Li P, Venkatayogi S (2018) Hydrogeochemical evaluation of groundwater quality for drinking and irrigation purposes and integrated interpretation with water quality index studies. Environ Process 5:363–383

Ahankoub M, Ayati F, Abroud M (2022) Investigating groundwater status of Mal-e Khalifeh Plain in Chaharmahal and Bakhtiari Province. Iran J Environ Sci Stud 7:5240–5250

Ahmed J, Wong LP, Chua YP, Channa N (2020) Drinking water quality mapping using water quality index and geospatial analysis in primary schools of Pakistan. Water 12:3382

Akhtar N, Syakir Ishak MI, Bhawani SA, Umar K (2021) Various natural and anthropogenic factors responsible for water quality degradation: a review. Water 13:2660

Ali SA, Ahmad A (2020) Analysing water-borne diseases susceptibility in Kolkata Municipal Corporation using WQI and GIS based Kriging interpolation. GeoJournal 85:1151–1174

Arab Amiri M, Mesgari MS (2019) Spatial variability analysis of precipitation and its concentration in Chaharmahal and Bakhtiari province, Iran. Theor Appl Climatol 137:2905–2914

Aryafar A, Khosravi V, Zarepourfard H, Rooki R (2019) Evolving genetic programming and other AI-based models for estimating groundwater quality parameters of the Khezri plain, Eastern Iran. Environ Earth Sci 78:1–13

Asghari FB, Mohammadi AA, Dehghani MH, Yousefi M (2018) Data on assessment of groundwater quality with application of ArcGIS in Zanjan, Iran. Data Brief 18:375–379

Avand M, Janizadeh S, Naghibi SA et al (2019) A comparative assessment of random forest and k-nearest neighbor classifiers for gully erosion susceptibility mapping. Water 11:2076

Avand M, Khiavi AN, Khazaei M, Tiefenbacher JP (2021) Determination of flood probability and prioritization of sub-watersheds: a comparison of game theory to machine learning. J Environ Manag 295:113040. https://doi.org/10.1016/j.jenvman.2021.113040

Badeenezhad A, Tabatabaee HR, Nikbakht H-A et al (2020) Estimation of the groundwater quality index and investigation of the affecting factors their changes in Shiraz drinking groundwater, Iran. Groundw Sustain Dev 11:100435

Balinski M, Laraki R (2007) A theory of measuring, electing, and ranking. Proc Natl Acad Sci USA 104:8720–8725. https://doi.org/10.1073/pnas.0702634104

Betrie GD, Tesfamariam S, Morin KA, Sadiq R (2013) Predicting copper concentrations in acid mine drainage: a comparative analysis of five machine learning techniques. Environ Monit Assess 185:4171–4182

Bhunia GS, Keshavarzi A, Shit PK et al (2018) Evaluation of groundwater quality and its suitability for drinking and irrigation using GIS and geostatistics techniques in semiarid region of Neyshabur, Iran. Appl Water Sci 8:1–16

Botsis D, Latinopulos P, Diamantaras K (2011) Rainfall-runoff modeling using support vector regression and artificial neural networks. In: 12th International conference on environmental science and technology (CEST2011), pp 8–10

Breiman L (2001) Random forests. Mach Learn 45:5–32

Burri NM, Weatherl R, Moeck C, Schirmer M (2019) A review of threats to groundwater quality in the anthropocene. Sci Total Environ 684:136–154

Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20:273–297

Davraz A, Özdemir A (2014) Groundwater quality assessment and its suitability in Çeltikçi plain (Burdur/Turkey). Environ Earth Sci 72:1167–1190

Demir S, Sahin EK (2022) Comparison of tree-based machine learning algorithms for predicting liquefaction potential using canonical correlation forest, rotation forest, and random forest based on CPT data. Soil Dyn Earthq Eng 154:107130

Ebrahime Moghadam F, Abbasnejad A (2020) Hydrogeochemical evaluation, groundwater quality and arsenic concentration of Sirjan plain using GIS and AqQa softwares. J Env Geo 14:1–24

Eid MH, Elbagory M, Tamma AA et al (2023) Evaluation of groundwater quality for irrigation in deep aquifers using multiple graphical and indexing approaches supported with machine learning models and GIS techniques, Souf Valley, Algeria. Water 15:182

Eissa MA, Mahmoud HH, Shouakar-Stash O et al (2016) Geophysical and geochemical studies to delineate seawater intrusion in Bagoush area, Northwestern coast, Egypt. J Afr Earth Sci 121:365–381

El Asri H, Larabi A, Faouzi M (2019) Climate change projections in the Ghis-Nekkor region of Morocco and potential impact on groundwater recharge. Theor Appl Climatol 138:713–727

Elkind E, Lang J, Saffidine A (2011) Choosing collectively optimal sets of alternatives based on the condorcet criterion. IJCAI Int Jt Conf Artif Intell. https://doi.org/10.5591/978-1-57735-516-8/IJCAI11-042

El-Magd SAA, Amer RA, Embaby A (2020) Multi-criteria decision-making for the analysis of flash floods: a case study of Awlad Toq-Sherq, Southeast Sohag, Egypt. J Afr Earth Sci 162:103709

El-Magd SAA, Pradhan B, Alamri A (2021) Machine learning algorithm for flash flood prediction mapping in Wadi El-Laqeita and surroundings, Central Eastern Desert, Egypt. Arab J Geosci 14:1–14

El-Magd SAA, Ahmed H, Pham QB et al (2022) Possible factors driving groundwater quality and its vulnerability to land use, floods, and droughts using hydrochemical analysis and GIS approaches. Water 14:4073

Elsayed S, Ibrahim H, Hussein H et al (2021) Assessment of water quality in Lake Qaroun using ground-based remote sensing data and artificial neural networks. Water 13:3094

Esfandiari F, Ghorbani Filabadi R, Nasiri Khiavi A, Mostafazadeh R (2019) Assessing the accuracy of algebraic and geostatistical techniques to determine the spatial variations of groundwater quality in Boroojen Plain. J Nat Environ Hazards 8:115–130

Fernández-Delgado M, Sirsat MS, Cernadas E et al (2019) An extensive experimental survey of regression methods. Neural Netw 111:11–34. https://doi.org/10.1016/j.neunet.2018.12.010

Frank E, Trigg L, Holmes G, Witten IH (2000) Naive Bayes for regression. Mach Learn 41:5–25

Gharechaee HR, Nazari Samani A, Sigaroodi K et al (2022) Groundwater salinity risk assessment in the southern plains of the Bakhtegan watershed using statistical and data mining models and fuzzy hierarchical analysis process. Watershed Manag Res J 35:2–14

Gholamrazai R (2020) An assessment of the groundwater quality using the AqQA model and determination of the most suitable method for their zoning (case study: Rafsanjan City, The Province of Kerman). J Environ Sci Technol 6:1–14

Hamlat A, Guidoum A (2018) Assessment of groundwater quality in a semiarid region of Northwestern Algeria using water quality index (WQI). Appl Water Sci 8:220

He S, Wu J, Wang D, He X (2022) Predictive modeling of groundwater nitrate pollution and evaluating its main impact factors using random forest. Chemosphere 290:133388

Heshmati S, Beigi H (2012) Spatial variability and mapping of shahrekord groundwater quality for use in the design of irrigation systems. J Water Res Agric 26:43–59

Hong H, Pradhan B, Xu C, Bui DT (2015) Spatial prediction of landslide hazard at the Yihuang area (China) using two-class kernel logistic regression, alternating decision tree and support vector machines. CATENA 133:266–281

Houéménou H, Tweed S, Dobigny G et al (2020) Degradation of groundwater quality in expanding cities in West Africa. A case study of the unregulated shallow aquifer in Cotonou. J Hydrol 582:124438

Ilić M, Srdjević Z, Srdjević B (2022) Water quality prediction based on Naive Bayes algorithm. Water Sci Technol 85:1027–1039

Iqbal J, Nazzal Y, Howari F et al (2018) Hydrochemical processes determining the groundwater quality for irrigation use in an arid environment: the case of Liwa Aquifer, Abu Dhabi, United Arab Emirates. Groundw Sustain Dev 7:212–219

Jamshidzadeh Z, Barzi MT (2018) Groundwater quality assessment using the potability water quality index (PWQI): a case in the Kashan plain, Central Iran. Environ Earth Sci 77:1–13

Jehan S, Khan S, Khattak SA et al (2019) Hydrochemical properties of drinking water and their sources apportionment of pollution in Bajaur agency, Pakistan. Measurement 139:249–257

Karthik K, Mayildurai R, Mahalakshmi R et al (2019) Physicochemical analysis of groundwater quality of Velliangadu area in Coimbatore District, Tamilnadu, India. Rasayan J Chem 12:409–414

Khan MA, Khan N, Ahmad A et al (2023) Potential health risk assessment, spatio-temporal hydrochemistry and groundwater quality of Yamuna river basin, Northern India. Chemosphere 311:136880

Khiavi AN, Vafakhah M, Sadeghi SH (2022) Comparative prioritization of sub-watersheds based on flood generation potential using physical, hydrological and co-managerial approaches. Water Resour Manag 36:1897–1917

Koranga M, Pant P, Kumar T et al (2022) Efficient water quality prediction models based on machine learning algorithms for Nainital Lake, Uttarakhand. Mater Today Proc 57:1706–1712

Kotsiantis SB, Pintelas PE (2004) Combining bagging and boosting. Comput Intell 1:324–333. https://doi.org/10.1103/PhysRevD.77.085025

Krishna Kumar S, Logeshkumaran A, Magesh NS et al (2015) Hydrogeochemistry and application of water quality index (WQI) for groundwater quality assessment, Anna Nagar, part of Chennai City, Tamil Nadu, India. Appl Water Sci 5:335–343

Kubicz J, LochyńskiPaweładPawełczyk A, Karczewski M (2021) Effects of drought on environmental health risk posed by groundwater contamination. Chemosphere 263:128145

Kumar A, Malyan SK, Kumar SS et al (2019a) An assessment of trace element contamination in groundwater aquifers of Saharanpur, Western Uttar Pradesh, India. Biocatal Agric Biotechnol 20:101213

Kumar R, Kumar R, Prakash O (2019b) Chapter-5 the impact of chemical fertilizers on our environment and ecosystem. Chief Ed 35:69

Lashkari H, Esfandiari N, Kashani A (2021) European journal of climate change. J Ref Eur J Clim Ch 3:20–32

Lee S-M, Park K-D, Kim I-K (2020) Comparison of machine learning algorithms for Chl-a prediction in the middle of Nakdong River (focusing on water quality and quantity factors). J Korean Soc Water Wastewater 34:277–288

Leong WC, Bahadori A, Zhang J, Ahmad Z (2021) Prediction of water quality index (WQI) using support vector machine (SVM) and least square-support vector machine (LS-SVM). Int J River Basin Manag 19:149–156

Li P, He S, Yang N, Xiang G (2018) Groundwater quality assessment for domestic and agricultural purposes in Yan'an City, northwest China: implications to sustainable groundwater quality management on the Loess Plateau. Environ Earth Sci 77:1–16

Lianjun C (2016) Research on snow extracting methods on the basis of random forests algorithm. Int J Simul Syst Sci Technol 17:3.1–3.6. https://doi.org/10.5013/IJSSST.a.17.19.03

Madani K (2010) Game theory and water resources. J Hydrol 381:225–238
Maghrebi M, Noori R, Bhattarai R et al (2020) Iran's agriculture in the anthropocene. Earth's Futur 8:e2020EF001547
Mahjouri N, Bizhani-Manzar M (2013) Waste load allocation in rivers using fallback bargaining. Water Resour Manag 27:2125–2136
Mallick J, Singh CK, AlMesfer MK et al (2018) Hydro-geochemical assessment of groundwater quality in Aseer Region, Saudi Arabia. Water 10:1847
Masoud M, El Osta M, Alqarawy A et al (2022) Evaluation of groundwater quality for agricultural under different conditions using water quality indices, partial least squares regression models, and GIS approaches. Appl Water Sci 12:244
Mirzaei A, Saghafian B, Mirchi A, Madani K (2019) The groundwater–energy–food nexus in Iran's agricultural sector: implications for water security. Water 11:1835
Mohamed S, El-Sabrouty MN et al (2014) Applications of hydrogeochemical modeling to assessment geochemical evolution of the Quaternary aquifer system in Belbies area, East Nile Delta, Egypt. J Biol Earth Sci 4:E34–E47
Morgenstern U, Daughney CJ (2012) Groundwater age for identification of baseline groundwater quality and impacts of land-use intensification—the National Groundwater Monitoring Programme of New Zealand. J Hydrol 456:79–93
Mousavi A, Solaimani K, Shokrian F, Roshun SH (2020) Investigation of spatio-temporal variation in groundwater resource quality using geo-statistical methods (case study: Lordegan Plain, Chaharmahal and Bakhteyari Province). Irrig Water Eng 10:262–275
Nafouanti MB, Li J, Mustapha NA et al (2021) Prediction on the fluoride contamination in groundwater at the Datong Basin, Northern China: comparison of random forest, logistic regression and artificial neural network. Appl Geochem 132:105054
Naghibi SA, Vafakhah M, Hashemi H et al (2020) Water resources management through flood spreading project suitability mapping using frequency ratio, k-nearest neighbours, and random forest algorithms. Nat Resour Res 29:1915–1933
Noi PT, Degener J, Kappas M (2017) Comparison of multiple linear regression, cubist regression, and random forest algorithms to estimate daily air surface temperature from dynamic combinations of MODIS LST data. Remote Sens. https://doi.org/10.3390/rs9050398
Noori R, Ghiasi B, Salehi S et al (2022) An efficient data driven-based model for prediction of the total sediment load in rivers. Hydrology 9:36
Norouzi H, Moghaddam AA (2020) Groundwater quality assessment using random forest method based on groundwater quality indices (case study: Miandoab plain aquifer, NW of Iran). Arab J Geosci 13:1–13
Ojekunle ZO, Adeyemi AA, Taiwo AM et al (2020) Assessment of physicochemical characteristics of groundwater within selected industrial areas in Ogun State, Nigeria. Environ Pollut Bioavailab 32:100–113
Osisanwo FY, Akinsola JET, Awodele O et al (2017) Supervised machine learning algorithms: classification and comparison. Int J Comput Trends Technol 48:128–138
Padarian J, McBratney AB, Minasny B (2020) Game theory interpretation of digital soil mapping convolutional neural networks. Soil 6:389–397
Panaskar DB, Wagh VM, Muley AA et al (2016) Evaluating groundwater suitability for the domestic, irrigation, and industrial purposes in Nanded Tehsil, Maharashtra, India, using GIS and statistics. Arab J Geosci 9:1–16
Pei-Yue L, Hui Q, Jian-Hua W (2011) Application of set pair analysis method based on entropy weight in groundwater quality assessment—a case study in Dongsheng City, Northwest China. E-J Chem 8:851–858
Peters J, Verhoest NEC, Samson R et al (2008) Wetland vegetation distribution modelling for the identification of constraining environmental variables. Landsc Ecol 23:1049–1065
Pham LT, Luo L, Finley A (2021) Evaluation of random forests for short-term daily streamflow forecasting in rainfall—and snowmelt-driven watersheds. Hydrol Earth Syst Sci 25:2997–3015. https://doi.org/10.5194/hess-25-2997-2021
Rahimi D (2012) Potential ground water resources: (case study: Shahrekord plain). Geo and Env plan 4:127–142
Rahman MM, Karunasinghe J, Clifford S et al (2020) New insights into the spatial distribution of particle number concentrations by applying non-parametric land use regression modelling. Sci Total Environ 702:134708
Ramadan FSM (2016) Sedimentological and hydrogeochemical studies of the quaternary groundwater aquifer in El Salhyia area, Sharkia governorate, Egypt. Middle East J Appl Sci
Ramadhani D, Afdal M, Rahmawita M et al (2021) The classification status of river water quality in riau province using modified K-nearest neighbor algorithm with STORET modeling and water pollution index. J Phys Conf Ser 6:120–138
Rao NS, Rao PS, Reddy GV et al (2012) Chemical characteristics of groundwater and assessment of groundwater quality in Varaha River Basin, Visakhapatnam District, Andhra Pradesh, India. Environ Monit Assess 184:5189–5214
Rashid A, Farooqi A, Gao X et al (2020) Geochemical modeling, source apportionment, health risk exposure and control of higher fluoride in groundwater of sub-district Dargai, Pakistan. Chemosphere 243:125409
Rasool U, Yin X, Xu Z et al (2022) Mapping of groundwater productivity potential with machine learning algorithms: a case study in the provincial capital of Baluchistan, Pakistan. Chemosphere 303:135265
Ravish S, Setia B, Deswal S (2021) Groundwater quality in urban and rural areas of north-eastern Haryana (India): a review. ISH J Hydraul Eng 27:224–234
Rawat KS, Singh SK, Gautam SK (2018) Assessment of groundwater quality for irrigation use: a peninsular case study. Appl Water Sci 8:1–24
Ren J, Lee SD, Chen X et al (2009) Naive bayes classification of uncertain data. In: 2009 ninth IEEE international conference on data mining, pp 944–949
Rozos E (2019) Machine learning, urbanwater resources management and operating policy. Resources. https://doi.org/10.3390/RESOURCES8040173
Safa G, Najiba C, El Houda BN et al (2020) Assessment of urban groundwater vulnerability in arid areas: case of Sidi Bouzid aquifer (central Tunisia). J Afr Earth Sci 168:103849
Sahour S, Khanbeyki M, Gholami V et al (2023) Evaluation of machine learning algorithms for groundwater quality modeling. Environ Sci Pollut Res 30:1–18
Salehi H, Soleymani L, Ebrahimi Mohammadi S (2016) An assessment of the groundwater quality using the AqQA model and determination of the most suitable method for their zoning (case study: Qoryeh City, The Province of Kurdistan). Wat Res Eng 29:30–49
Sarath Prasanth SV, Magesh NS, Jitheshlal KV et al (2012) Evaluation of groundwater quality and its suitability for drinking and agricultural use in the coastal stretch of Alappuzha District, Kerala, India. Appl Water Sci 2:165–175
Sarker B, Keya KN, Mahir FI et al (2021) Surface and ground water pollution: causes and effects of urbanization and industrialization in South Asia. Guigoz Sci Rev 7:32–41
Shabani S, Yousefi P, Adamowski J, Naser G (2016) Intelligent soft computing models in water demand forecasting. Water Stress Plants. https://doi.org/10.5772/63675

Sharifi Garmdareh E, Vafakhah M, Eslamian SS (2018) Regional flood frequency analysis using support vector regression in arid and semi-arid regions of Iran. Hydrol Sci J 63:426–440. https://doi.org/10.1080/02626667.2018.1432056

Sharma MK, Kumar M (2020) Sulphate contamination in groundwater and its remediation: an overview. Environ Monit Assess 192:1–10

Siebert S, Burke J, Faures J-M et al (2010) Groundwater use for irrigation—a global inventory. Hydrol Earth Syst Sci 14:1863–1880

Singh MK, Jha D, Jadoun J (2012) Assessment of physico-chemical status of groundwater samples of Dholpur District, Rajasthan. India Int J Chem 4:96

Singh AK, Raj B, Tiwari AK, Mahato MK (2013) Evaluation of hydrogeochemical processes and groundwater quality in the Jhansi district of Bundelkhand region, India. Environ Earth Sci 70:1225–1247

Srdjevic Z, Bajcetic R, Srdjevic B (2012) Identifying the criteria set for multi-criteria decision making based on SWOT/PESTLE analysis: a case study of reconstructing a water intake structure. Water Resour Manag 26:3379–3393. https://doi.org/10.1007/s11269-012-0077-2

Starzyk J (2010) Water resource planning and management using motivated machine learning. IAHS-AISH Publ 338:214–220

Subramani T, Elango L, Damodarasamy SR (2005) Groundwater quality and its suitability for drinking and agricultural use in Chithar River Basin, Tamil Nadu, India. Environ Geol 47:1099–1110

Suvarna B, Reddy YS, Sunitha V et al (2020) Data on application of water quality index method for appraisal of water quality in around cement industrial corridor, Yerraguntla Mandal, YSR District, AP South India. Data Brief 28:104872

Talebiniya M, Khosravi H, Zohrabi S (2019) Assessing the ground water quality for pressurized irrigation systems in Kerman Province, Iran using GIS. Sustain Water Resour Manag 5:1335–1344

Tao H, Hameed MM, Marhoon HA et al (2022) Groundwater level prediction using machine learning models: a comprehensive review. Neurocomputing 489:71–308

Tesoriero AJ, Gronberg JA, Juckem PF et al (2017) Predicting redox-sensitive contaminant concentrations in groundwater using random forest classification. Water Resour Res 53:7316–7331

Tien Bui D, Shirzadi A, Shahabi H et al (2019) A novel ensemble artificial intelligence approach for gully erosion mapping in a semi-arid watershed (Iran). Sensors 19:2444

Tiwari AK, Singh AK (2014) Hydrogeochemical investigation and groundwater quality assessment of Pratapgarh district, Uttar Pradesh. J Geol Soc India 83:329–343

Tsangaratos P, Ilia I (2016) Comparison of a logistic regression and Naive Bayes classifier in landslide susceptibility assessments: the influence of models complexity and training dataset size. CATENA 145:164–179

Tyagi S, Sarma K (2020) Qualitative assessment, geochemical characterization and corrosion-scaling potential of groundwater resources in Ghaziabad district of Uttar Pradesh, India. Groundw Sustain Dev 10:100370

Udmale P, Ichikawa Y, Manandhar S et al (2014) Farmers′ perception of drought impacts, local adaptation and administrative mitigation measures in Maharashtra State, India. Int J Disaster Risk Reduct 10:250–269

Vafakhah M, KhosrobeigiBozchaloei S (2020) Regional analysis of flow duration curves through support vector regression. Water Resour Manag 34:283–294. https://doi.org/10.1007/s11269-019-02445-y

Vafakhah M, Nasiri Khiavi A, Janizadeh S, Ganjkhanlo H (2022) Evaluating different machine learning algorithms for snow water equivalent prediction. Earth Sci Inform 15:2431–2445

Vorpahl P, Elsenbeer H, Märker M, Schröder B (2012) How can statistical models help to determine driving factors of landslides? Ecol Model 239:27–39

Wang F, Wang Y, Zhang K et al (2021) Spatial heterogeneity modeling of water quality based on random forest regression and model interpretation. Environ Res 202:111660

Wu J, Xue C, Tian R, Wang S (2017) Lake water quality assessment: a case study of Shahu Lake in the semiarid loess area of northwest China. Environ Earth Sci 76:1–15

Xie Z, Lou I, Ung WK, Mok KM (2012) Freshwater algal bloom prediction by support vector machine in Macau storage reservoirs. Math Probl Eng. https://doi.org/10.1155/2012/397473

Zakaria N, Anornu G, Adomako D et al (2021) Evolution of groundwater hydrogeochemistry and assessment of groundwater quality in the Anayari catchment. Groundw Sustain Dev 12:100489

Zamani-Ahmadmahmoodi R, Alimirzaee Z, Gharahi N, Najafi M (2019) The relationship of land use and quality of groundwater resources of Chaharmahal and Bakhtiari Province. J Nat Environ 72:353–364

Zhang Q, Qian H, Xu P et al (2021) Groundwater quality assessment using a new integrated-weight water quality index (IWQI) and driver analysis in the Jiaokou Irrigation District, China. Ecotoxicol Environ Saf 212:111992. https://doi.org/10.1016/j.ecoenv.2021.111992

Zhou J, Li E, Wei H et al (2019) Random forests and cubist algorithms for predicting shear strengths of rockfill materials. Appl Sci 9:1–16. https://doi.org/10.3390/app9081621