Journal of Cleaner Production 360 (2022) 131847

Contents lists available at ScienceDirect

Journal of Cleaner Production

journal homepage: www.elsevier.com/locate/jclepro

Use of machine learning and geographical information system to predict nitrate concentration in an unconfined aquifer in Iran

V. Gholami [a,*], M.J. Booij [b]

[a] Department of Range and Watershed Management and Dept. of Water Eng. and Environment, Faculty of Natural Resources, University of Guilan, Sowmeh Sara, 1144, Guilan, Iran
[b] Water Engineering and Management, Faculty of Engineering Technology, University of Twente, the Netherlands

ARTICLE INFO

Handling Editor: Bin Chen

Keywords:
Extreme gradient boosting
Deep neural network
Multiple linear regression
Nitrate pollution
Alluvial aquifer.

ABSTRACT

Increased nitrate concentration is one of the main groundwater quality problems today that needs to be measured and monitored. Water quality testing and monitoring are time consuming and costly. Therefore, new modeling methods such as machine learning algorithms can be used as an efficient solution for predicting nitrate concentration. In this study, three machine learning methods including deep neural network (DNN), extreme gradient boosting (EGB), and multiple linear regression (MLR) were used to predict nitrate contamination in groundwater in the north of Iran (Mazandaran plain) and finally the best method was selected for mapping. The mean nitrate concentration in 250 piezometric wells was considered as output variable and the factors affecting groundwater quality (groundwater depth, transmissivity of aquifers, precipitation, evaporation, distance from water resources and Caspian Sea, distance from industries and residential centers, population density, topography, and exploitation from groundwater) as input variables in an alluvial aquifer. The same training and testing data were used in the modeling process of the three machine learning methods. The results of the training and testing stages showed that the EGB method has the highest performance in predicting nitrate concentration due to the lowest error values and highest correlation between the measured and predicted values of nitrate concentration (training R-sqr = 0.98, Nash–Sutcliffe efficiency (NSE) = 0.98, and test R-sqr = 0.86, NSE = 0.84). Further, the results indicate that the factors distance from industries, population density, groundwater depth, and evaporation rates are the most important factors affecting nitrate concentration in groundwater. Finally, the tested EGB model and a geographic information system (GIS) tool were used to prepare a map of groundwater nitrate pollution in the study area. Evaluating the performance of the resulting map by comparing the predicted and measured values indicated a good accuracy (R-sqr = 0.8).

1. Introduction

Nitrate concentration in groundwater has increased significantly in recent decades due to human activities, especially with the development of industries and the use of nitrogen fertilizers in agriculture (Sajedi-Hosseini et al., 2018; Motevalli et al., 2019; Band et al., 2020; Khalifa et al., 2021). Industrial and agricultural development without complying environmental standards, especially in developing countries such as Iran, has led to increased pollution of surface water and groundwater resources (Shivasorupy et al., 2012; Gholami et al., 2015). Industries in developing countries do not have wastewater treatment systems or environmental standards required. Farmers also traditionally use excessive amounts of nitrogen fertilizers and toxins. As a result, increasing nitrate concentrations in groundwater sources have been observed in recent decades in the world (Alighardashi and Mehrani, 2017; Band et al., 2020). On the other hand, consuming drinking water with high nitrate concentration can lead to serious problems for human health, especially for children such as childhood water syndrome and lack of oxygen in the body., high concentrations of nitrate in drinking water will increase the risk of cancer for humans (Parvizishad et al., 2017; Taneja et al., 2017). Therefore, it is necessary to conduct monitoring and modeling studies to identify the factors affecting the increase of nitrate concentration, identify areas with high pollution, and ways to reduce it.

* Corresponding author. Department of Range and Watershed Management and Dept. of Water Eng. and Environment, Faculty of Natural Resources, University of Guilan, Sowmeh Sara, 1144, Guilan, Iran.
E-mail address: Gholami.vahid@guilan.ac.ir (V. Gholami).

https://doi.org/10.1016/j.jclepro.2022.131847
Received 6 December 2021; Received in revised form 14 April 2022; Accepted 16 April 2022
Available online 12 May 2022

Sampling groundwater and conducting water quality measurements are time consuming and costly. On the other hand, developing countries face financial limitations regarding water quality tests, and monitoring studies, and mostly such studies are not the priority of development projects or can not be implemented on a large scale. Therefore, using new methods and modeling techniques to predict nitrate concentration and contamination zones can be an effective way under such conditions. New effective methods in modeling the quality of water resources are the use of artificial intelligence and machine learning methods. Today, several studies on the quality and quantity of water resources have been conducted worldwide using artificial intelligence and machine learning methods. Artificial intelligence in groundwater quality studies (Chou, 2006; Han et al., 2011; Band et al., 2020; Gholami et al., 2020; Maliqi et al., 2020; Mosaffa et al., 2021) and studies on groundwater depth fluctuations (Dixon, 2004; Saemi and Ahmadi, 2008; Gong et al., 2018; Chen et al., 2020; Gholami et al., 2021) has been widely used. Further, machine learning was used in different hydrological modeling studies with a high performance (Rahmati et al., 2017; Tongal and Booij, 2018; Rahmati et al., 2019; Azizi et al., 2020; Kashani et al., 2020; Javidan and Javidan, 2021; Wells et al., 2021).

Wang and ZhangDing (2017) used machine learning and a water quality index (WQI) for modeling groundwater quality in China. They found good results in the test stage (R-sqr of 0.92) that showed a high performance of machine learning in the water quality modeling. Motevalli et al. (2019) used data mining methods and a geographic information system (GIS) to investigate nitrate contamination in groundwater in the Ghaemshahr plain. They were able to provide an accurate map for nitrate concentration zoning. Rahmati et al. (2019) studied uncertainty of machine learning methods for predicting nitrate pollution in groundwater using UNEEC methods and quantile regression. They used three state-of-the-art ML models including support vector machine (SVM), random forest (RF), and k-nearest neighbor (KNN) and found that KNN was the best model among the used models. Bedi et al. (2020) compared the performance of three methods for modeling groundwater quality. They used artificial neural network (ANN), SVM, and EGB. They observed the highest correlation between the observed and the predicted values and the lowest errors in the modeling process by the EGB model. Awais et al. (2021) used several machine learning approach for evaluating nitrate contamination risks along the Karakoram Highway. They used SVM, multivariate siscriminant analysis (MDA), and boosted regression trees (BRT) in the modeling process and their results showed that machine learning have a good performance in nitrate pollution evaluation. Bilali et al. (2021) used different methods of machine learning in groundwater quality prediction. Their results showed that selecting a suitable method for modeling water quality will reduce costs and can evaluate water quality in a short time. Khalifa et al. (2021) predicted groundwater nitrate concentration using artificial intelligence methods in a semi-arid region. They found that the factors precipitation, altitude, groundwater depth, and distance from the residential area are the most important factors for predicting nitrate pollution. Further, many studies have shown a high performance of combining artificial intelligence and GIS capabilities in groundwater quality modeling (Tweed et al., 2007; Arslan, 2012; Haselbeck et al., 2019; Abulibdeh et al., 2021).

There is an urgent need to study and monitor the rates of nitrate pollution in groundwater for the management and use of water resources. It is also necessary to provide a model with the ability to predict the effect of different scenarios of agricultural and industrial development and population growth on nitrate pollution in groundwater. The aim of this study is therefore to present a methodology for predicting groundwater nitrate concentration and to prepare a groundwater nitrate concentration zoning map by combining the capabilities of machine learning and GIS techniques and spatial variation in groundwater of an alluvial aquifer on the southern coats of the Caspian Sea. Furthermore, the application of different machine learning methods and their evaluation in order to determine the most effective method for predicting the concentration of nitrate in groundwater is investigated.

2. Materials and methods

2.1. Study area

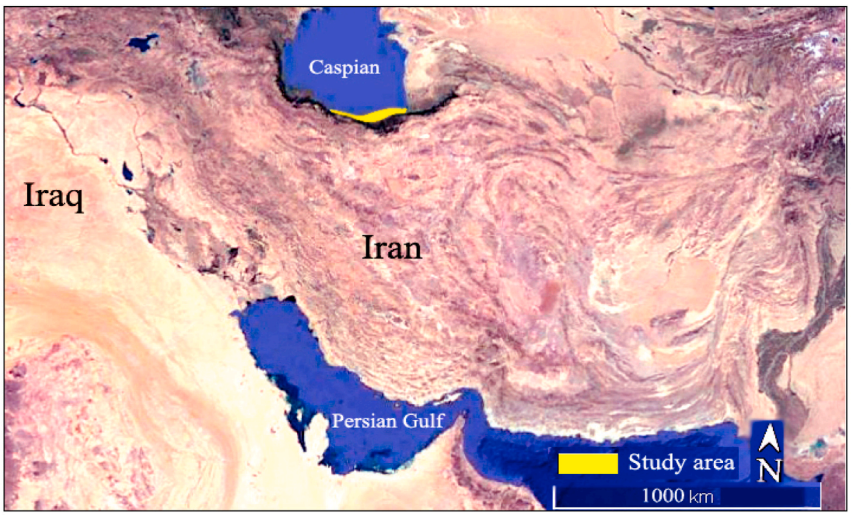
The study area includes the Mazandaran plain with an area of ten thousand km^2 in the north of Iran. This plain includes the southern coasts of the Caspian Sea located between 50°34′ to 54°10′ eastern longitude and 35°47′ to 37° northern latitude (Fig. 1). The mean annual precipitation of the study area varies from 1300 mm in the west of the plain to 600 mm in the east of the plain. The mean annual evaporation range from 600 to 1000 mm. Land uses include agricultural lands, residential lands (cities and villages), water resources, industrial lands, and a limited area of forest land (Fig. 2E). Quaternary alluvial sediments are the main constituent of the plain formation. An unconfined aquifer (alluvial aquifer) with a shallow groundwater layer and rivers are the main sources of water supply for agricultural lands in the study area. The groundwater table varies from 1 m below surface level at the coast to 40 m below surface level in the most elevated part.

The activity of a large number of small industries without complying environmental standards in the study plain threatens the groundwater quality. In addition to the activities of these industries, the excessive use of fertilizers in agriculture, in particular, irrigated agriculture has reduced the quality of water resources and especially increased nitrate concentration in recent decades (Alighardashi and Mehrani, 2017).

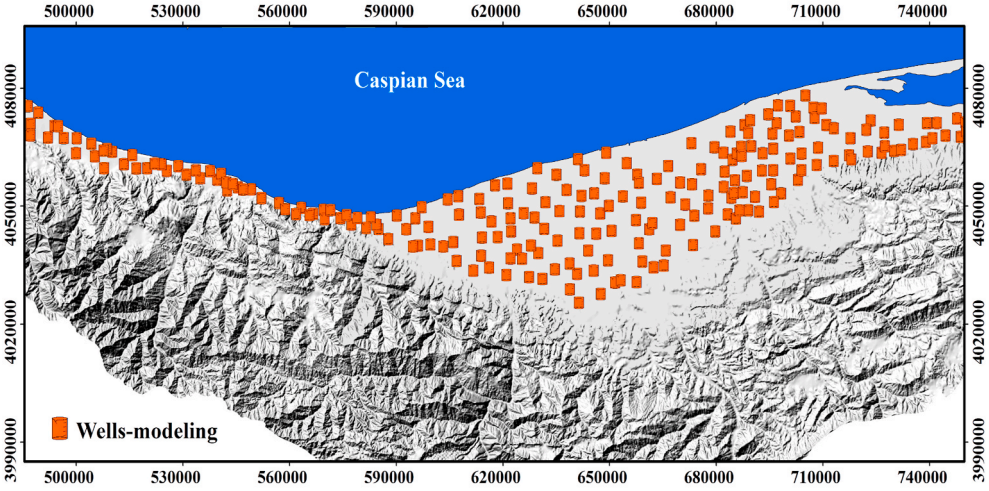
2.2. Data

For modeling the nitrate concentration in groundwater, accurate data (measured data) on the concentration values in groundwater are needed, as well as observations of the affecting factors for the nitrate concentration. Therefore, data on nitrate concentration in groundwater of 250 piezometric wells in the Mazandaran plain was obtained from the Mazandaran Regional Water Company (Mazandaran Regional Water Company MRWC, 2020). Further, ten pumping wells were selected and sampled in the eastern part of the study plain and their measured values of nitrate concentration were used to compare and evaluate the modeling results (verification wells). The verification wells are located in areas without samples for the model (training and testing).

The quantitive data on nitrate concentration was used for the years 2016–2020. Each well had a monthly sampling frequency. The piezometric wells were semi-deep wells (20–40 m) drilled in the alluvial aquifer. Nitrate pollution in unconfined aquifers is higher due to the extent of paddy lands, the use of nitrogen fertilizers and the activities of large and small industries (Alighardashi and Mehrani, 2017). There are deep wells or drinking water wells in the study area with a depth of more than 100 m, which have less pollution. The piezometric wells were drilled by the MRWC to study the fluctuations of the groundwater depth, prepare water level maps and calculate the volume of groundwater inlet and outlet of the aquifer. On the other hand, the number of deep wells in the study plain is limited and is not sufficient for modeling, so deep wells have not been used in combination or separately. Finally, the average of all monthly observations of nitrate concentration for each well was estimated and used as output variable in the modeling process (training and testing phases). Affecting factors for the nitrate concentration in the Mazandaran plain have been studied using hydrogeologic maps of MRWC, and satellite images (Google Earth). For each of the 250 wells, the affecting factors for the nitrate concentration in groundwater include groundwater depth, transmissivity of aquifers, climate (precipitation and evaporation), distance from industries, distance from water sources, population density, distance from residential centers, distance from the sea, topography, and exploitation from groundwater (according to landuse and groundwater extractions in wells), and these factors were carefully investigated and determined. Nitrogen fertilizer consumption is an important factor for nitrate pollution in groundwater.



(A)



(B)

Fig. 1. (A) Location of the Mazandaran plain in the north of Iran and (B) location of the studied wells across the Mazandaran plain.

Unfortunately, no data is available on the amount of fertilizer consumption and only the mean consumption per capita in agricultural lands has been estimated, which is considered as a criterion of consumption of nitrogen fertilizers by studying land use within the operating radius of each well.

2.2.1. Groundwater depth

Groundwater depth is one of the influecting factors for groundwater quality (Ghose et al., 2010; Bradai et al., 2016; Haselbeck et al., 2019; Li et al., 2020). The data on groundwater fluctuations from the last five years, which were almost normal years (no severe drought), were used to determine the mean groundwater depth (Fig. 2A). Then, the mean groundwater depth map was prepared using data of 250 piezometric wells and interpolation using Kriging in GIS. The map shows the spatial variability of the groundwater depth in the study area.

The depth of the well below the water table is one of the effective factors for the prediction of nitrate pollution in groundwater. Unfortunately, detailed data on this factor is not available for the study area, and this limits the application of the model for nitrate concentration mapping. Therefore, this factor has not been used as an input factor.

Well depth can also be an influecting factor for the nitrate concentration in groundwater (Bohlke, 2002). The purpose of this study is to provide a model for predicting nitrate concentration at locations without data or wells, so well depth could not be used as an input to the model because it limits the use of the model. Further, based on the available data, there is no significant relationship between well depth and nitrate concentration in groundwater in the study area.

2.2.2. Climate (precipitation and evaporation)

Precipitation is the main source of recharge for surface water resources and finally groundwater.

In agricultural areas, more precipitation can cause a large loss of nitrate to groundwater. Further, precipitation can reduce the concentration of pollutants and improve water quality. Therefore, the role of precipitation in nitrate concentration in determining groundwater is depend on hydrological conditions of the aquifer, farm management, and climate conditions (Band et al., 2020; Khalifa et al., 2021). Further, evaporation can also increase the concentration of pollutants in water resources (Awasthi ea tal., 2005; Maliqi et al., 2020). Annual precipitation and actual evaporation data of meteorological stations were prepared for a common 30-year period (1990–2020). Then, by interpolating the stations statistics, the spatial distribution of mean annual

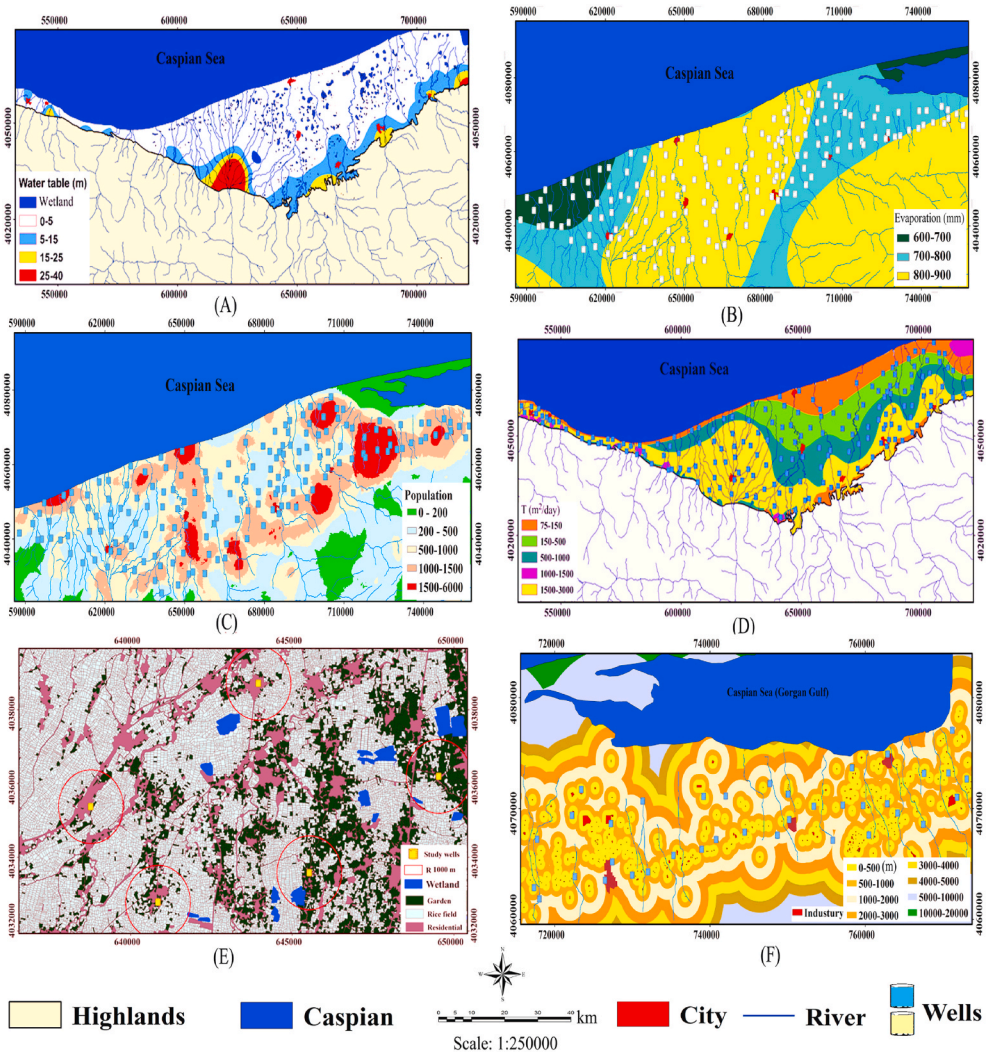


Fig. 2. Maps of (A) Mean depth to groundwater (m), rivers, and wetlands across the study plain; (B) Mean annual evaporation (mm); (C) Population density (no/km2); (D) Mean transmissivity of aquifer formation (m2/day); (E) Land use map for calculating exploitation values from groundwater; and (F) Distance from industries (m).

precipitation and evaporation (Fig. 2B) in the study area was determined.

2.2.3. Population density and distance from residential centers

Residential areas are important zones that contaminate groundwater resources (Khalifa et al., 2021). The study area, a considerable number of cities and villages is located, where all villages do not have any new and efficient sewage system. Furthermore, small industries are located in these cities and their suburbs. As a result, these centers are an important factor in polluting water resources and increasing nitrate concentrations (Gholami et al., 2015). Using topographic maps with a scale of 1: 25,000 and satellite images, the digital layer of villages and cities was prepared in GIS (Fig. 2E) and then the map with the distance from residential centers was prepared (Fig. 2F). Moreover, by interpolating the population statistics of the residential centers, the spatial distribution of population and population density in the study area was determined and mapped (Fig. 2C).

2.2.4. Transmissivity of aquifers

The transmissivity of aquifers is one of the most important hydrogeological characteristics of an aquifer and has a significant effect on the quality of groundwater resources and the spread of pollutants (Awasthi et al., 2005; Gholami et al., 2021). The transmissivity of the aquifer has been determined by MRWC by drilling and pumping tests. Then, based on hydrogeological studies and geological maps, homogeneous units of transmissivity were determined. Taking into account these homogeneous units and the estimated values in the wells, a map of the transmissivity of the aquifer formation has been prepared by MRWC. The map shows the mean values of the transmissivity in square meters per day (Mazandaran Regional Water Company MRWC, 2020) (Fig. 2D).

2.2.5. Distance from industries

Industries are one of the most important factors influencing the concentration of nitrate in groundwater resources and increasing pollution of water resources (Shivasorupy et al., 2012In the Mazandaran plain, industrial towns, small industries, livestock and poultry farms have a significant effect on the pollution of water resources and increase in nitrate concentration (Gholami et al., 2015). To determine the location of industries, we have used the database of the Mazandaran Province Industries and Mines Organization (MPIMO) and visual interpretation of satellite images with a high resolution. After preparing the geo-referenced digital layer of industries, a map of the industrial area of the study plain was prepared in GIS (Fig. 2F).

2.2.6. Distance from water sources

The distance from water sources such as rivers and lakes of fresh

water and wetlands will affect the amount of recharge, the groundwater depth and finally the quantity and quality of groundwater (Awasthi et al., 2005; Gholami et al., 2021). Using topographic maps and satellite images, freshwater resources of the plain, including rivers, wetlands and lakes (Fig. 2B&E) were identified and a buffer map (map of the distance from rivers, lakes and wetlands) was prepared in GIS the seam as Fig. 2 F.

2.2.7. Distance from the sea

Distance from the Caspian Sea can be an important factor for water quality as well, especially for groundwater salinity. The distance from the sea shows the location of the place in the watershed, which in turn will affect the groundwater depth and the amount of groundwater recharge (Abd-Elhamid et al., 2020; Shi et al., 2021). The Caspian Sea location and GIS capabilities were used to map the distance from the sea (Fig. 2).

2.2.8. Topography (elevation and slope)

Elevation and slope are important topographic factors influencing groundwater conditions (Khalifa et al., 2021). The elevation and slope of the land affect the groundwater depth, the drainage conditions and the hydraulic slope of the area (Wang and ZhangDing, 2017). At this stage, using 10 m topographic lines and GIS capabilities, a 10 m digital elevation model (DEM) was prepared and used to prepare a slope map of the study area. The ground slope is less than 5 percent and does not show notable variability in the study area.

2.2.9. Exploitation from groundwater

Agricultural lands, especially paddy fields and citrus orchards, are very important in consuming pollutants such as nitrogen fertilizers and toxins (Band et al., 2020; Costantini et al., 2021; Wells et al., 2021). The areas of agricultural lands also indicate water consumption in the agricultural sector. On average, 1 ha of paddy land consumes about 11,000 m^3 per year and 1 ha of citrus orchard consumes about 5,000 m^3 of water per year in the study plain (Mazandaran Regional Water Company MRWC, 2020). The area of agricultural lands in the study plain was determined by using high resolution images (Google Earth) and by applying the annual water consumption within the operating radius of the well. The maximum operating radius of the pumping well was considered to be 1000 m (Shi et al., 2016). The operating radius of the well is the maximum distance from the well up to which aquifer properties have a significant influence on drawdown at the well. By applying the annual water consumption of agricultural lands, the amounts of water utilization within the operating radius of each well were estimated. The area of agricultural land also reflects the amount of nitrogen fertilizer consumption because there is a direct relationship between the area of agricultural land and fertilizer consumption (Fig. 2E). The predominant crop is rice, which has a high nitrate fertilizer consumption. In the central part of the plain, rice is cultivated twice a year by farmers, which has led to excessive use of nitrogen fertilizers. Farmers in the region traditionally consume about 150–250 kg of nitrogen fertilizer per hectare per year (Mazandaran Regional Water Company MRWC, 2020).

2.3. Predicting the nitrate concentration in groundwater

To model the nitrate concentration in groundwater, three machine learning methods including deep neural network (DNN), extreme gradient boosting (EGB), and multiple linear regression (MLR) using the same training and test data have been used. In this regard, nitrate concentrations in groundwater of monitoring wells were used as output and factors affecting in nitrate concentration as input. 70% of the data was used for model training and 30% for model testing. The modeling process was performed by the three methods and then the results were compared including evaluation of error to determine the most appropriate method. Furthermore, the optimal inputs of the models have been determined through sensitivity analysis of the inputs in the model, trial and error method, and evaluation of the effect of each input and statistical analyzes. Finally, for each of the three used methods, the optimal model structure was determined. Then, the most efficient method was selected by comparing the results of the training and testing phases.

2.3.1. Extreme gradient boosting (EGB) method

The EGB method using decision trees is one of the most efficient methods in modeling with machine learning. The decision tree algorithm uses a tree-like model that goes from input variable data (branches of the tree) to the modeling of the output variable (leaves of the tree). This method has several advantages including no restrictions on the different types of output, ability to model complex interactions, and to manage missing data with minimal data missed. The EGB method was first presented by Breiman et al. (1984) and Mason et al. (2000); Chen and Guestrin (2016) have developed the EGB method. In this method a set of inputs $(X_1, \ldots, X_n)$ is used to model a set of outputs $(Y_1, \ldots, Y_n)$ by using a model $F(X) \rightarrow Y$ and to minimize the sum of loss function J by optimizing the model F(X) (eq. (1)):

$$J = \sum_{i=1}^{n} L(Yi, F(Xi)) \quad (1)$$

The loss function is $L(x, y) = (x - y)^2$. Model optimization algorithm firstly, computes the negative gradients of J with respect F(Xi) $(-\frac{\partial J}{\partial F(X_i)})$. Next, the regression tree h is fitted to the negative gradients $-\frac{\partial J}{\partial F(X_i)}$. Then, the new F(Xi) will be $F(Xi) + \gamma h$, where γ is called the step size in the algorithm to assess the calculated minimum of J, and this process is repeated until an appropriate accuracy is achieved. One of the salient features of the EGB method is that the modeling process begins with a loss function $L(Y_i, F(X_i) + h)$ and minimizes $J = \sum_{i=1}^{n} L(Y_i, F(X_i) + h) + \Omega(h)$, where $\Omega(h) = \gamma T + \frac{1}{2}\lambda||W||^2$ T is called the leaf number in the tree, and w is called the leaf weights. In this study, the R program was applied for modeling nitrate concentration in the groundwater by using the EGB model. A schematic diagram of the EGB model is given in Fig. 3.

2.3.2. Deep neural network (DNN) method

The DNN method is another machine learning methods which mainly uses nonlinear relationships to relate input data and output variables in the modeling process. A DNN has a structure similar to the neurons in the human brain that enables activity and communication between the inputs to train and learn the DNN and ultimately to model a target or output variable (Hu et al., 2018). The difference between DNN and neural networks is that neural networks use three categories (input layer, hidden layer and output layer) and DNN uses multiple hidden layers for the modeling process.

In this study, a multilayer perceptron (MLP) was used to develop an adopted DNN method. MLP is a category of feed forward artificial neural networks. The base of an MLP network is a perceptron network (Gholami et al., 2019). The perceptron has one or more inputs, a transfer function, and an output. A supervised learning algorithm of binary classifiers is used in the perceptron algorithm. The binary classifier is a function which can define whether or not the inputs, introduced by a vector of numbers, belongs to some specific class. Further, back propagation is the most commonly used algorithm for supervised training in multilayered neural networks that changes the nonlinear relationship between the inputs and the output by changing the weight values internally (Hu et al., 2018). The back propagation process was performed into two stages of feed forward and back propagation. In the feed forward method, a pattern is considered for the inputs and the effect of the pattern is spread from layer to layer through the network until the output is obtained. In the next step, the output values are compared with the observed values and an error signal is sent to each of the output neurons. Output errors are propagated backwards by the hidden layer. This process goes step by step until each neuron in the network receives an error signal, and this error signal represents the relative share of the total error. The goal is to achieve a minimum error function. The output

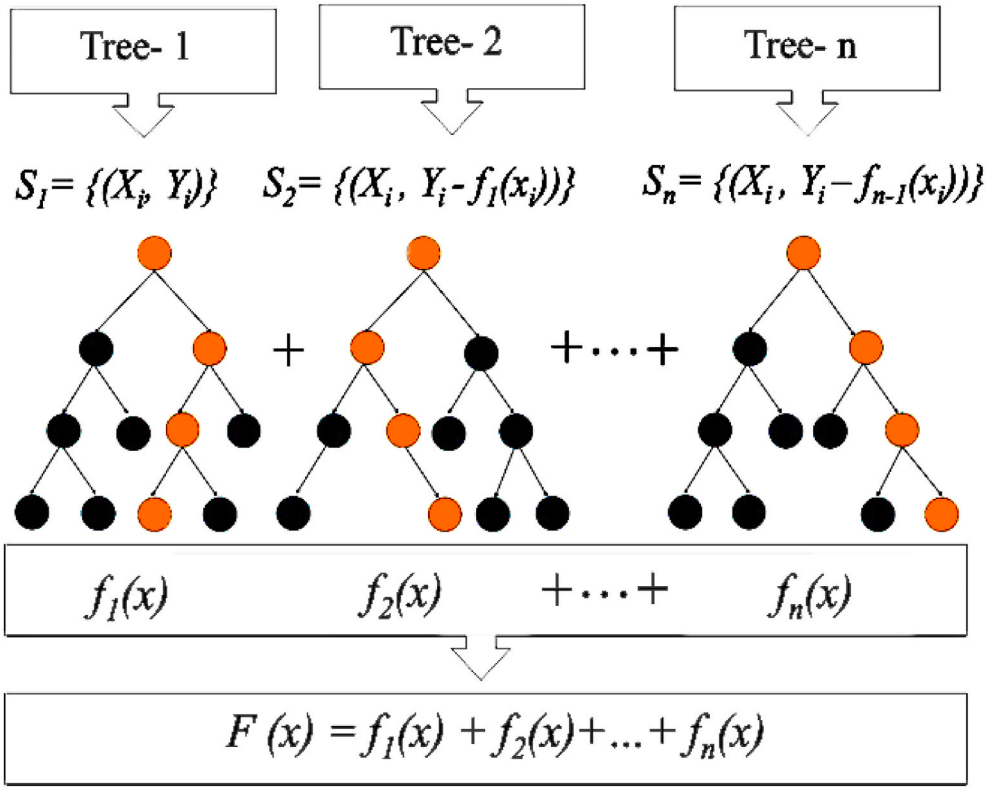


Fig. 3. A schematic diagram of the EGB approach in modeling nitrate concentration in groundwater.

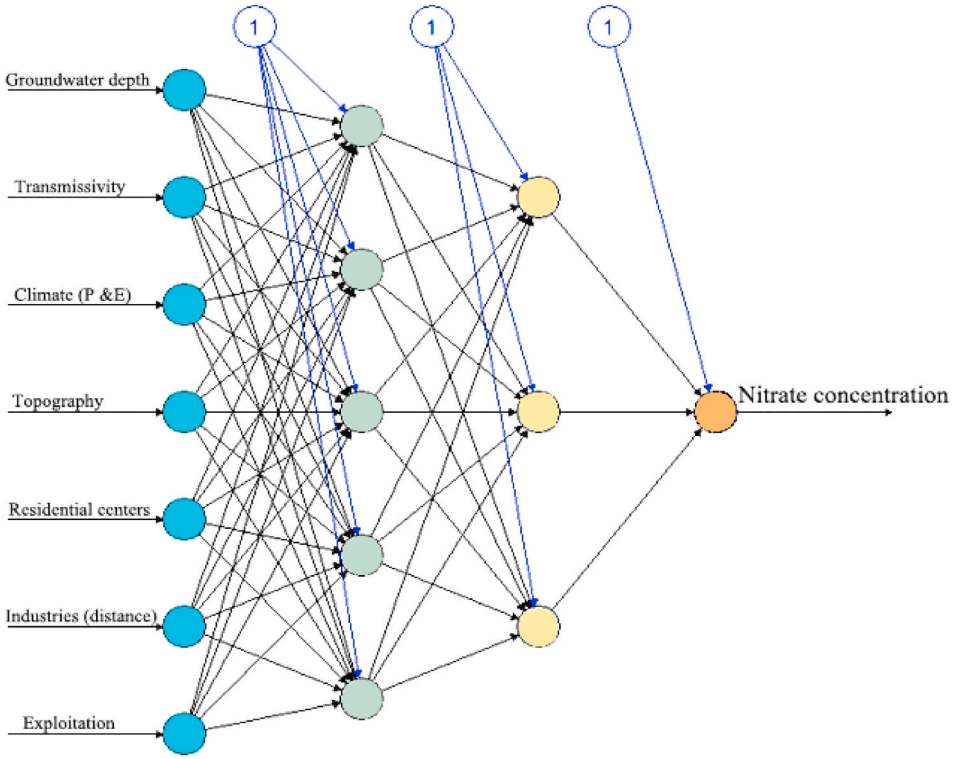


Fig. 4. The DNN structure for nitrate concentration prediction including input layers, hidden layers, and one output layer. The blue and black lines indicate the connections with weights and bias terms, respectively.

of a neuron in the hidden layer is obtained from the following equation:

$$H_j = f\left(\sum_{i=1}^{n} W_{ji}X_i + b_i\right) \quad (2)$$

where W_{ji} is the weights of the hidden layer neurons, b_i is the biases of the hidden layer neurons, and $f(.)$ is a nonlinear activation function. Further, the network output is estimated by equation (3):

$$y = f\left(\sum_{j=1}^{m} W_{kj}H_i + b_0\right) \quad (3)$$

where W_{kj} is the weights of the output layer neuron, b_0 is the biases of the output layer neuron, and $f(.)$ is the activation function of the output layer neuron (Aggarwal, 2018).

In this study, the Neuralnet package in R was used for application of the DNN in the modeling of nitrate concentration in the groundwater (Fritsch et al., 2019). In the modeling process, a trial- and-error method was used until the predicted and observed values of nitrate concentration achieved the best fit. The evaluation of the model performance was assessed using several error criteria. The optimal structure of the used DNN is determined based on minimum error values and maximum performance in predicting nitrate concentrations. The obtained optimal structure includes four input layers, two hidden layers, and one output layer. These two hidden layers have eight hidden neurons according to Fig. 4 The number of epochs indicates the number of complete repetitions of the training data set. The optimal number of epochs in the modeling process was 135.

2.3.3. Multiple linear regression (MLR) method

The MLR method derives patterns in the data and establishes the best fitting linear relationships between two or more independent variables and the target variable (nitrate concentration in groundwater). A stepwise approach was used in which the selection of variables is carried out by entering or removing input variables using criteria such as the coefficient of determination, the F-test, and the *t*-test. The regression method for *n* input variables $X_1, X_2, \ldots, X_n$ can be explained as follows. In a MLR model, each value of the input variable X is related to the value of the target variable Y. The linear regression relation is as follows:

$$Y = B_0 + B_1X_1 + B_2X_2 + \ldots + B_nX_n \quad (4)$$

Where Y is the simulated value of the target variable, B_0 is the Y value when all input variables are equal to zero, $X_1, \ldots, X_n$ are the inputs (affecting factors of the nitrate concentration in groundwater), and B_1 through B_n are the regression coefficients (Valentini et al., 2021). One of the features of the MLR method is the multi-linearity of the relations, i.e. the conditions in which two or more input variables are correlated with each other, and this is evaluated by using the inflation coefficient of variance (VIF). The multi-linearity of the model provides the ability to investigate the effect of single inputs on the output variable and finally to select the optimal input variables for the model. Using VIF is one of the most widely used methods for determining multi-linearity (Reiser, 2004). Distance from Caspian sea and site elevation have multi-collinearity (VIF >11). The training process is performed to reach the maximum agreement between the predicted and observed values. Basically, linear models are easier to use than nonlinear methods and black box models. However, for values with a high variance, i.e. maximum and minimum values, they have a lower efficiency and larger errors than nonlinear methods (Gholami et al., 2021).

2.4. Performance evaluation of the models

The performance evaluation of the used models was carried out through the comparison between predicted and measured values in the testing stage. Several statistical criteria were used to evaluate the model performances such as the R-squared values, the Nash-Sutcliffe efficiency (NSE) and the normalized root mean squared deviation (NRMSD) as follows:

$$RMSD = \sqrt{\sum_{i=1}^{n} \frac{\left(Y_o - \widehat{Y}_p\right)^2}{n}} \quad (5)$$

$$NSE = 1 - \frac{\sum_{i=1}^{n}\left(\widehat{Y}_p - Y_o\right)^2}{\sum_{i=1}^{n}\left(Y_o - \overline{Y}_{oi}\right)^2} \quad (6)$$

$$NRMSD = \frac{RMSD}{\overline{Y}_{oi}} \quad (7)$$

where Y_o is the measured nitrate concentration value, is the predicted nitrate concentration value, n is the number of samples, and is the mean of the measured nitrate concentration values. Moreover, feature importance (FI) was used to analyze the strength of the relationship between input variables and target variable (nitrate concentration). The estimation of the feature importance is based split variable of regression tree in the iteration (Friedman and Meulman, 2003). The relative FI was estimated for all of inputs and has a range between 0 and 1.

2.5. Mapping the nitrate concentration in groundwater

The optimal inputs of the final model for estimating the nitrate concentration as raster layers were prepared in GIS as a raster layers. Then, by coupling the capabilities of the machine learning method and GIS, the nitrate concentration values were predicted in the entire area of Mazandaran plain. Finally, the nitrate concentration map of groundwater was prepared. The accuracy of the map was evaluated by comparing the measured values of nitrate concentration in the 250 piezometric wells and the predicted values on the map. Ten independent pumping wells were selected and sampled in the study plain and their measured values of nitrate concentration were compared with the prepared map as verification data. Sampling and measurement of the verification wells have been done in areas without samples for the training and testing stages.

3. Results

The observed mean nitrate concentration in groundwater in the 250 piezometric wells was determined between 0.7 and 113.3 mg/l with a mean concentration in Mazandaran plain of 23.4 mg/l. According to the Iranian drinking water standard, the maximum concentration of nitrate in drinking water is 50 mg/l. Therefore, groundwater in 35 of the 250 studied wells had nitrate concentrations exceeding the Iranian drinking water standard. The maximum concentration of nitrate of 220 mg/l in summer and the minimum value of 0.1 mg/l in winter were observed in two different wells. In summer, industry activities and agricultural activities (nitrate fertilizer consumption) increase in the study area.

Groundwater depth data showed that the mean annual depth of groundwater in the study plain varies between 1 and 40 m. In addition, the transmissivity of aquifer formations varies between 50 and 3,500 m^2/day based on past studies (Mazandaran Regional Water Company MRWC, 2020; Gholami et al., 2021) and existing hydrogeologic maps. Minimum values are observed in heavy-textured (fine-grained) formations and maximum values are found in light-textured (coarse-grained) formations. The minimum elevation at the coasts of the Caspian Sea is −27 m below mean sea level and the maximum elevation is about 100 m on the border of the plain and the highlands area. In terms of distance from water sources and residential centers, the multiplicity of rivers, wetlands, lakes and residential centers has caused the distance of study wells to vary from a few meters to several kilometers. Annual precipitation varies in the plain between 600 and 1300 mm. The maximum precipitation in the western part of the study plain is due to the proximity of the sea and mountains and the effect of the Caspian Sea. Annual

evaporation varies between 600 and 1000 mm per year, the maximum of which is observed in the eastern part of the plain due to higher temperatures. The population density in the area is highly variable and its amount varies between 0 and 6000 people per square kilometer. The highest values are observed in cities and lowest values in forests or agricultural lands. Based on agricultural areas and exploitation values in active exploitation wells in the operating radius of wells (1000 m), annual exploitation values were estimated between 0.1 and 3.66 million m^3 per year.

Statistical analysis was performed between inputs data (factors affecting the nitrate concentration in groundwater) and model output (nitrate concentration in groundwater). The correlation coefficients between nitrate concentrations and input data show involvement of input factors in nitrate pollution in groundwater and can be found in Table 1. Moreover, to determine the optimal inputs and to determine the effective factors influencing the nitrate concentration in groundwater, a sensitivity analysis of inputs in the used models has been carried out (Fig. 5). The results in Fig. 5 show that factors such as distance from industry, population density, evaporation and groundwater depth have the greatest effect on variations in nitrate concentration in groundwater in the study plain. Other studies showed similar results (Shivasorupy et al., 2012; Gholami et al., 2015; Alighardashi and Mehrani, 2017). The FI analysis showed similar results in determining the main inputs for nitrate concentration in groundwater.

On the other hand, in the modeling process, a trial and error method was used to determine the optimal inputs showing that the factors distance from industry, population density, groundwater depth, and evaporation are the optimal inputs in modeling nitrate pollution. The model training process used in all three methods showed good results, which are presented in Table 2. The inputs of all three models were exactly the same. Based on the results, the EGB and the DNN methods show an acceptable performance in estimating the nitrate concentration in groundwater. However, in the training phase, the EGB method was more efficient and had lower error rates compared to the other two methods. After training the models, testing was performed and the results of the test phase of the three methods are presented in Table 2 as well. Based on the results and evaluation of the indices values, the EGB method is the most efficient model to predict the values of nitrate concentration in groundwater. Further, comparison between the predicted and measured values in the training and test stages of the three models is presented in Fig. 6. Based on the results, the EGB model has the highest correlation (R-squred in Table 2 and Fig. 6) between the predicted and measured values in the training and test phases.

Finally, the tested EGB model was used as the most efficient method to predict the values of nitrate concentration in groundwater in the entire area of Mazandaran plain. The maps of the optimal inputs of the model including distance from industries, annual evaporation, groundwater depth and exploitation values in GIS were prepared. Finally, we used these inputs and tested the EGB model to predict the nitrate concentration in groundwater in the Mazandaran plain as shown in Fig. 7. Furthermore, two methods were used to evaluate the performance of the methodology and the accuracy of the results. First, the measured values of nitrate concentration in groundwater in 250 wells used in the modeling process (training and testing) were compared with the predicted values. According to the results, a high accuracy was observed (R-sqr = 0.8). Second, the nitrate concentration values in groundwater were measured in 10 pumping wells (Fig. 7) other than the wells used for training and testing (east of the study plain) and used as verification wells. Therefore, the observed values of nitrate concentration in these wells were compared with the values of the map of nitrate concentration. In the verification stage, the R-sqr between the predicted values and the measured values was equal to 0.79.

The model used in predicting nitrate concentration shows a good performance and the resulting map has the appropriate accuracy to be used as a source of information for planning and management for groundwater resources. It is important in the modeling process of nitrate concentration to identify places with excessive rates of nitrate concentration (>50 mg/l) to take the necessary measures and planning to reduce the concentration and manage associated risks and adverse effects.

Table 1
Pearson's correlation coefficients between the nitrate concentrations of groundwater in the 250 piezometric wells, and the input factors (nitrate concentration is in mg/l).

Input factor	Correlation with nitrate concentration	Significance (P-value)
Groundwater depth	−0.15	0.01
Transmissivity of aquifer	0.1	0.03
Elevation	0.07	0.49
Slope	0.02	0.37
Annual evaporation	0.15	0.01
Distance from water resources	0.09	0.06
Distance from Caspian sea	0.14	0.01
Annual precipitation	−0.08	0.08
Population density	0.56	0.00
Distance from residential areas	−0.06	0.1
Distance from industries	−0.57	0.00
Groundwater exploitation	0.14	0.00
Well depth	−0.06	0.1

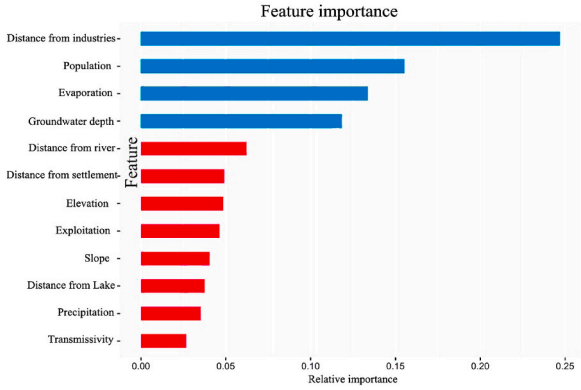


Fig. 5. Sensitivity analysis of the input variables of the optimum model (EGB) using feature importance (FI). Factors with a larger effect on nitrate concentration in groundwater are marked in blue and less important factors are marked in red. Factors are arranged in order of influence on nitrate concentration from top to bottom.

Table 2
Performance of the three models in the training and test stages for modeling nitrate concentration in groundwater.

	NSE		NRMSD		R-squared	
	Training	Test	Training	Test	Training	Test
Extreme gradient boosting (EGB)	0.98	0.84	0.01	0.45	0.98	0.86
Deep neural network (DNN)	0.81	0.54	0.4	0.84	0.83	0.57
Multiple linear regression (MLR)	0.75	0.45	0.7	0.97	0.77	0.48

4. Discussion

Evaluation of measured nitrate concentrations and factors affecting nitrate concentration of groundwater in study wells showed that maximum values are mainly observed in industrial and residential centers (towns) with a high population density. Moreover, maximum

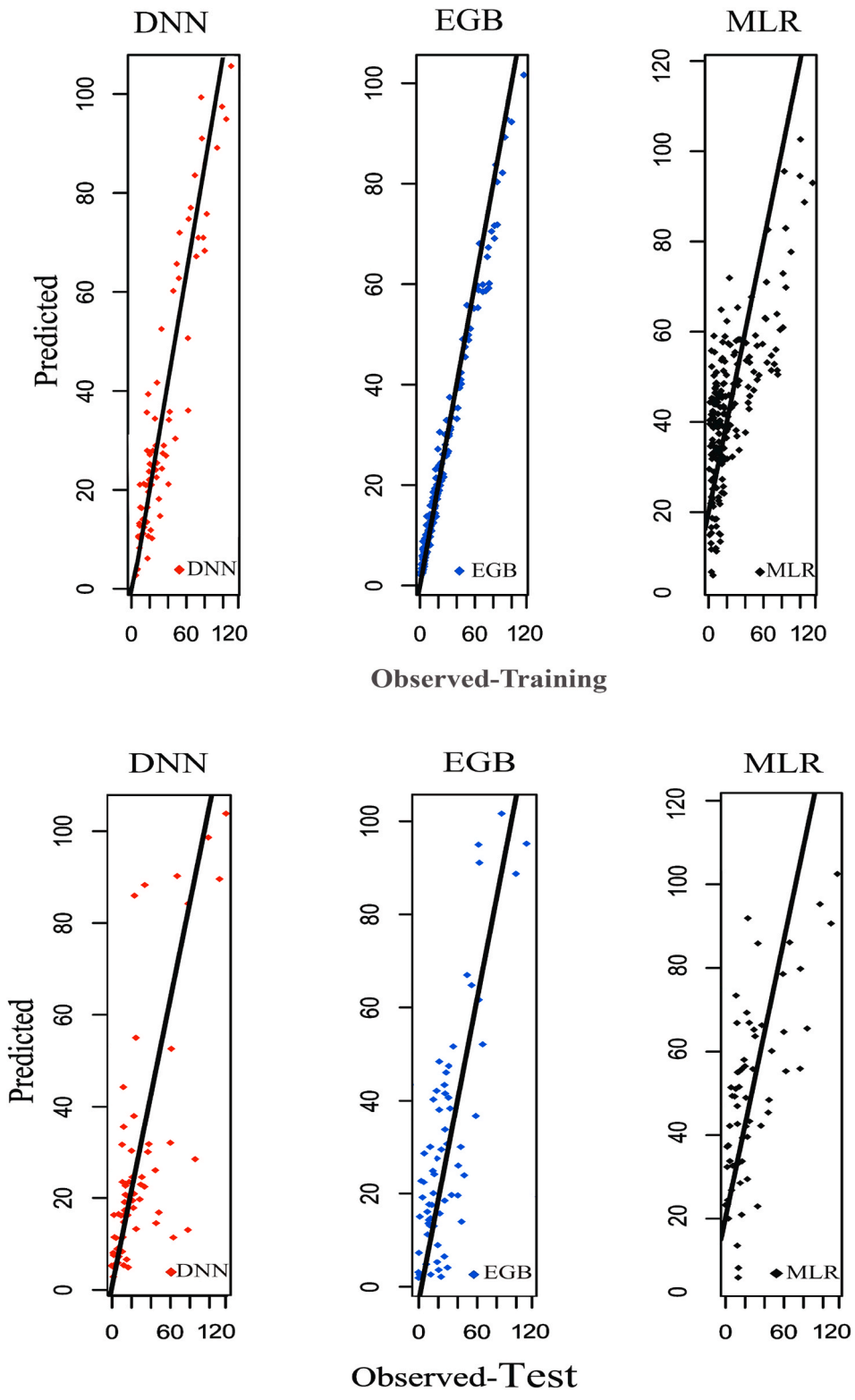


Fig. 6. The performance evaluation of the three models (EGB, DNN, and MLR) in the training and test phases of predicting nitrate concentration in groundwater (mg/l).

correlation coefficients were found between industrial and residential centers and nitrate concentration in groundwater. Therefore, the factors industrial and residential centers are the most important factors influencing groundwater nitrate concentration in Mazandaran plain. The distance from industries had the highest correlation (R = −0.57). The secondly most effective factor was the population density of residential centers (R = 0.56). Among the factors affecting groundwater quality that have been investigated in the present study, the factors of distance from industries, population density, annual evaporation values and groundwater depth have a significant and strong relationship with nitrate concentrations in groundwater (Nemcic-Jurec and Jazbec, 2017; Motevalli et al., 2019). Industrial centers, depending on the type of their

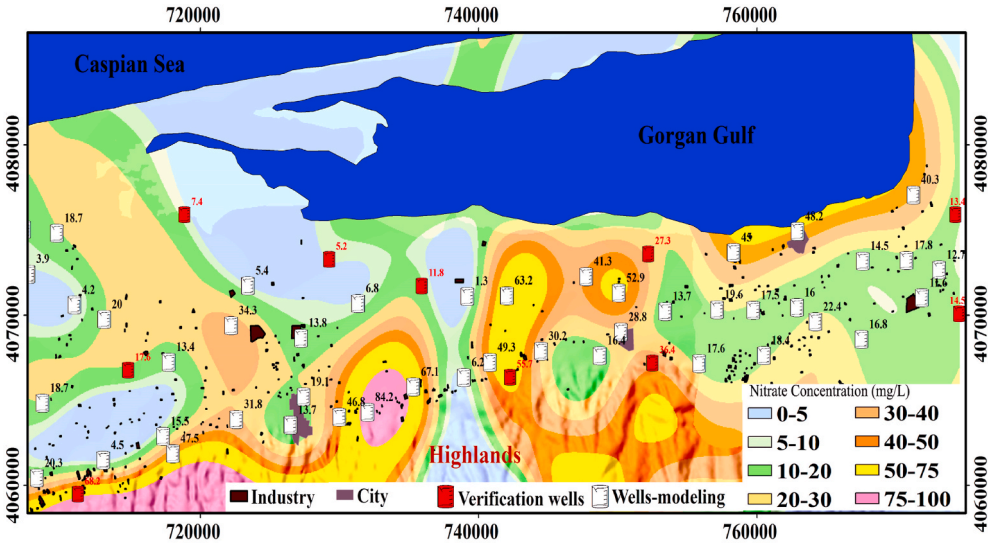


Fig. 7. Map of nitrate concentration in groundwater (mg/l) was generated using the outputs of the tested EGB model in GIS. The measured nitrate values in the groundwater (training, testing, and verification wells) fall within the predicted ranges on the map (to evaluate the accuracy of the results). The R-sqr, NSE, and NRMSD values for the verification wells were 0.79, 0.7, and 0.5, respectively.

products, have effluents and wastewaters with high amounts of nitrate (Band et al., 202; Khalifa et al., 2021). On the other hand, these industries in the study area often do not have standard wastewater treatment systems. In residential centers, human sewage, household wastewater, gardens and paddy fields, small livestock farms and cemeteries are important factors in increasing nitrate concentration (Wells et al., 2021). The higher the population density in residential centers, the larger the number of activities and the more effluent and sewage will be produced.

Another important factor is the depth of the groundwater. Groundwater depth is inversely related to nitrate concentration values (FI = 0.12, R = −0.15), but does not have a strong relation with nitrate concentration. According to the results, places where the groundwater depth is less (water table is higher), the vulnerability of the groundwater and the spread of contamination is higher (Gholami et al., 2015; Nemcic-Jurec and Jazbec, 2017). Further, the depth of the wells, the location of the well screen, and the length of the well screen are influencing the vulnerability of the groundwater (Dash and SarangiSingh, 2010). Shallow aquifers have a higher potential for pollution, and the high water table in areas with paddy lands, orchards and industrial centers increase the conditions for the spread of pollution.

Another effective factor explaining the nitrate concentration is evaporation (R = 0.13). Evaporation has a significant, but weak relation with nitrate concentration. In places with more evaporation, temperature is higher and precipitation is less, and as a result, groundwater recharge will be lower. Further, more industries are observed in these areas of the study plain. As a result, nitrate concentration in groundwater is larger (Savard et al., 2007). Evaporation also indirectly affects the quality of water resources and increasing the concentration of pollutants in water resources (Awasthi et al., 2005; Maliqi et al., 2020).

Consumption of nitrogen fertilizers in agriculture lands is an important factor influencing nitrate pollution in groundwater. The development of agriculture and especially the second cultivation of rice in summer (twice per year) have increased the consumption of nitrogen fertilizers in the region, which has led to an increase in nitrate concentration. The predominant landuse of the study plain is paddy lands where agricultural lands have a greater role in non-point pollution and point nitrate pollution is more affected by existing industries (Motevalli et al., 2019). Unfortunately, no exact data are available on the use of nitrogen fertilizers in the study plain. If there is industrial activity in areas between paddy and residential lands, the values are much higher than the allowable nitrate concentration.

In the northern half of the study plain, due to high precipitation and significant river discharge, the groundwater depth is high (water table <5 m) and the risk of pollution is high (Gholami et al., 2015; Khalifa et al., 2021). In many areas of paddy lands, the groundwater depth varies between 1 and 5 m (shallow aquifer) and there is a high potential for vulnerability of groundwater resources in unconfined aquifers in this area (Nemcic-Jurec and Jazbec, 2017; Motevalli et al., 2019). The amount of groundwater recharge is very effective to improve the quality of groundwater. The groundwater quality after six month recharging during clod seasons (in the beginning of spring) is more proper and in the beginning of autumn (end of summer) low quality of groundwater will generally be observed (Qin et al., 2011). Other studied factors have a weak correlation or no significant relationship with the nitrate concentration (P value > 0.05).

The process of predicting nitrate concentrations in groundwater was performed using three machine learning methods. Based on the results, the performances of the three methods used varied. EGB model showed the highest performance and the MLR model had the lowest performance in modeling nitrate concentration values. Past research also indicates that the EGB model has a very high performance in comparison with other methods of machine learning and artificial intelligence in hydrological modeling for the prediction of water quantity and quality (Mason et al., 2000; Chen and Guestrin, 2016; Bedi et al., 2020). Machine learning is a subset of artificial intelligence that enables machines to learn without prior planning and using previous data and experience. In artificial intelligence, we design intelligent systems to do anything like humans and this includes has a wide range of modeling methods and requires high expertise and experience. Machine learning tries to build a model that can perform only the specific tasks for which it is trained.

Evaluation of test results by comparing error indices and comparing observational and predicted values indicate a high performance of the EGB model in the modeling process. However, the performance of the proposed models in predicting the minimum, average and maximum values of the mean monthly nitrate concentration in groundwater is not the same. The three used models generally predict average values with a high accuracy and show larger errors in predicting minimum and maximum values. The EGB model has the highest performance in predicting the minimum and maximum nitrate values compared to the other two methods. One of the reasons for this is that in the model training data we have mainly used minimum and average values, the

number of observed values of maximum nitrate concentration was limited. Such a problem is observed in many environmental modeling studies (Gholami et al., 2021). The importance of identifying and zoning areas with too high concentrations of nitrate in groundwater is well known. Therefore, to analyze and solve these problems, one should try to use wells or samples with maximum values in the modeling process, in particular in the training data.

In the modeling process, the data of the main inputs of the model (distance from industries, population density, evaporation, and groundwater depth) can be used to estimate nitrate concentration rates in groundwater in each place. The map of nitrate concentration can be an effective information source for water resources management as well as sharing the results of machine learning modeling for everyone. An important point about the models used to predict the nitrate concentration in groundwater is that the main inputs are the distance from industry, population density, evaporation, and groundwater depth, all of which are accessible or measurable.

Further, the inputs of industrial and residential lands and population density will change over time, which makes it possible to use the present models for temporal changes in nitrate concentration or to study the effect of different scenarios of industrial, agricultural and population development on nitrate pollution in groundwater.

5. Conclusion

Determining nitrate concentration and monitoring its changes is essential in water resources management. However, sampling and testing are time consuming and costly. Therefore, new modeling methods such as the use of machine learning methods can be used as an aid in the study of water resources quality. It should be noted that such modeling is not a substitute for sampling and testing but an addition to water quality studies to increase operating speed and reduce costs. In the modeling process, a suitable number of measured output data (nitrate concentration) is needed, and a higher number of samples will lead to a more accurate and efficient model.

In the modeling process with machine learning, three models were used. The goal is to obtain an efficient model that can be used in the study area with an effective number of main inputs. Based on the results of the present study, the EGB model has a very high ability to estimate the nitrate concentration values in groundwater. This model can be used by using GIS as a preprocessor and post-processor to predict spatial variations in nitrate concentration, monitor nitrate concentration, and finally identify areas with excessive concentrations. The first step in the modeling process of the nitrate concentration in groundwater is to determine the optimal inputs and accurately estimate them and subsequently the correct use of a model or machine learning algorithm. In the modeling process, we can present models with different inputs or up-to-date modeling methods with local conditions and available data in a short time and with high performance that can be connected to ther systems like GIS. The present modeling can also be used to predict spatial and temporal changes in nitrate concentrations in groundwater in other areas. In discussing temporal changes, different scenarios of population growth and industry development or other input factors can be used for the model and their effects on changes of nitrate concentration can be predicted using a tested model. For future studies, it is suggested that other input data and methods of artificial intelligence such as fuzzy neural networks, self-organizing maps or neural networks be used to predict nitrate concentration with a higher accuracy and for zoning of nitrate pollution in groundwater.

CRediT authorship contribution statement

V. Gholami: has been involved in field studies and nitrate concentration modeling process. **M.J. Booij:** has been involved in analyzing the results and writing the manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to thank the Mazandaran Regional Water Company (MRWC) for providing the hydrogeological data.

References

Abd-Elhamid, H.F., Abd-Elaty, I., Sherif, M.M., 2020. Effects of aquifer bed slope and sea level on saltwater intrusion in coastal aquifers. J. Hydrol. 7, 5. https://doi.org/10.3390/hydrology7010005, 2020.

Abulibdeh, A., Al-Awadhi, T., Nasiri, N., Buloshi, A., Abdelghani, M., 2021. Spatiotemporal mapping of groundwater salinity in Al-Batinah, Oman. Groundwater Sustain. Dev. 12, 100551 https://doi.org/10.1016/j.gsd.2021.100551, 2021.

Aggarwal, C.C., 2018. Neural Networks and Deep Learning, vol. 10. Springer, 978-3.

Alighardashi, A., Mehrani, M.J., 2017. Survey and zoning of nitrate-contaminated groundwater in Iran. J. Mater. Environ. Sci. 8 (12), 4339–4348, 2017.

Arslan, H., 2012. Spatial and temporal mapping of groundwater salinity using ordinary kriging and indicator kriging: the case of Bafra Plain. Turkey. Agric. Water Manag. 113, 57–63.

Awais, M., Aslam, B., Maqsoom, A., Khalil, U., Ullah, F., Azam, S., Imran, M., 2021. Assessing nitrate contamination risks in groundwater: a machine learning approach. Appl. Sci. 11, 10034. https://doi.org/10.3390/app112110034.

Awasthi, A.K., Dubey, O.P., Awasthi, A., Sharma, S., 2005. A Fuzzy Logic model for estimation of groundwater recharge. In: Annual Meeting of the North American Fuzzy Information Processing Society. Detroit, MI, June 26–28, 809–813.

Azizi, A., Gilandeh, Y.A., Mesri-Gundoshmian, T., Saleh-Bigdeli, A.A., Moghaddam, H.A., 2020. Classification of soil aggregates: a novel approach based on deep learning. Soil Tillage Res. 199, 104586 https://doi.org/10.1016/j.still.2020.104586.

Band, S.S., Janizadeh, S., Pal, S.C., Chowdhuri, I., Siabi, Z., Norouzi, A., Melesse, A.M., Shokri, M., Mosavi, A., 2020. Comparative analysis of artificial intelligence models for accurate estimation of groundwater nitrate concentration. Sensors 20, 5763. https://doi.org/10.3390/s20205763.

Bedi, S., Samal, A., Ray, C., Snow, D., 2020. Comparative evaluation of machine learning models for groundwater quality assessment. Environ. Monit. Assess. 192 https://doi.org/10.1007/s10661-020-08695-3.

Bilali, A., Taleb, A., Brouziyne, Y., 2021. Groundwater quality forecasting using machine learning algorithms for irrigation purposes. Agric. Water Manag. 245, 106625 https://doi.org/10.1016/j.agwat.2020.106625.

Bohlke, J.K., 2002. Groundwater recharge and agricultural contamination. Hydrogeol. J. 10, 153–617. https://doi.org/10.1007/s10040-001-0183-3, 179.

Bradai, A., Douaoui, A., Bettahar, N., Yahiaoui, I., 2016. Improving the prediction accuracy of groundwater salinity mapping using indicator kriging method. J. Irrigat. Drain. Eng. 142, 4016023.

Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984. Classification and Regression Trees. CRC press.

Chen, T., Guestrin, C., 2016. Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, pp. 785–794.

Chen, C., He, W., Zhou, H., Xue, Y., Zhu, M., 2020. A comparative study among machine learning and numerical models for simulating groundwater dynamics in the Heihe River Basin, northwestern China. Sci. Rep. 10, 3904. https://doi.org/10.1038/s41598-020-60698-9.

Chou, K., 2006. A review on integration of artificial intelligence into water quality modelling. Mar. Pollut. Bull. 52 (7), 726–733.

Costantini, M.L., Agah, H., Fiorentino, F., Irandoost, F., Leon Trujillo, F.J., Careddu, G., Calizza, E., Rossi, L., 2021. Nitrogen and metal pollution in the southern Caspian Sea: a multiple approach to bioassessment. Environ. Sci. Pollut. Res. 28, 9898–9912. https://doi.org/10.1007/s11356-020-11243-8, 2021.

Dash, J.P., Sarangi, A., Singh, D.K., 2010. Spatial variability of groundwater depth and quality parameters in the national capital territory of Delhi. J. Environ. Manag. 45, 640–650. https://doi.org/10.1007/s00267-010-9436-z, 2010.

Dixon, B., 2004. Prediction of groundwater vulnerability using an integrated GIS-based neuro-fuzy techniques. J. Spa. hydrol. 14 (12), 1–38.

Friedman, J.H., Meulman, J.J., 2003. Multiple additive regression trees with application in epidemiology. Stat. Med. 22, 1365–1381.

Fritsch, S., Guenther, F., Guenther, M.F., 2019. Package neuralnet. Train. Neural Network. 236 (1). Available online: ftp://64.50.

Gholami, V., Agha Goli, H., Kalteh, A.M., 2015. Modeling sanitary boundaries of drinking water wells on the Caspian Sea southern coasts, Iran. Environ. Earth Sci. 74 (4), 2981–2990.

Gholami, V., Torkman, J., Dalir, P., 2019. Simulation of precipitation time series using tree-rings, earlywood vessel features, and artificial neural network. Theor. Appl. Climatol. 137 (3), 1939–1948.

Gholami, V., Khaleghi, M.R., Taghvaye Salimi, E., 2020. Groundwater quality modeling using self-organizing map (SOM) and geographic information system (GIS) on the

Caspian southern coasts. J. Mt. Sci. 17, 1724–1734. https://doi.org/10.1007/s11629-019-5483-y, 2020.
Gholami, V., Khaleghi, M.R., Pirasteh, S., Booij, M.J., 2021. Comparison of self-organizing map, artificial neural network, and co-active neuro-fuzzy inference system methods in simulating groundwater quality: geospatial artificial. Water Resour. Manag. 36, 451–469. https://doi.org/10.1007/s11269-021-02969-2, 2022.
Ghose, D.K., Panda, S.S., Swain, P.C., 2010. Prediction of water table depth in western region, Orissa using BPNN and RBF neural networks. J. Hydrol. 394, 296–304.
Gong, Y.C., Wang, Z.J., Xu, G.Y., Zhang, Z., 2018. A Comparative study of groundwater level forecasting using data-driven models based on ensemble empirical mode decomposition. Water 10, 20. https://doi.org/10.3390/w10060730.
Han, H.G., Chen, Q.l., Qiao, J.F., 2011. An efficient self-organizing RBF neural network for water quality prediction. Neural Network. 24 (7), 717–725.
Haselbeck, V., Kordilla, J., Krause, F., Sauter, M., 2019. Self-organizing maps for the identification of groundwater salinity sources based on hydrochemical data. J. Hydrol. 576, 610–619.
Hu, C., Wu, Q., Li, H., Jian, S., Li, N., Lou, Z., 2018. Deep learning with a long short-term memory networks approach for rainfall-runoff simulation. Water 10 (11), 1543.
Javidan, R., Javidan, N., 2021. A novel artificial intelligence-based approach for mapping groundwater nitrate pollution in the Andimeshk-Dezful plain, Iran. Geocarto Int. https://doi.org/10.1080/10106049.2022.2035830.
Kashani, M.H., Ghorbani, M.A., Shahabi, M., Naganna, S.R., Diop, L., 2020. Multiple AI model integration strategy—application to saturated hydraulic conductivity prediction from easily available soil properties. Soil Tillage Res. 196, 104449 https://doi.org/10.1016/j.still.2019.104449.
Khalifa, M.A., Mukherjee, K., Pandey, M., Arora, A., Janizadeh, S., Pham, Q.B., Tran Anh, D., Ahmadi, K., 2021. Prediction of groundwater nitrate concentration in a semiarid region using hybrid Bayesian artificial intelligence approaches. Environ. Sci. Pollut. Res. (4) https://doi.org/10.1007/s11356-021-17224-9, 2021.
Li, J., Shi, Z., Liu, F., 2020. Evaluating spatiotemporal variations of groundwater quality in northeast Beijing by self-organizing map. Water 12 (5), 1382. https://doi.org/10.3390/w12051382.
Maliqi, E., Jusufi, K., Singh, S.K., 2020. Assessment and spatial mapping of groundwater quality parameters using metal pollution indices, graphical methods and geoinformatics. Anal. Chem. Lett. 10 (2), 152–180, 2020.
Mason, L., Baxter, J., Bartlett, P.L., Frean, M.R., 2000. Boosting algorithms as gradient descent. In: Advances in Neural Information Processing Systems, pp. 512–518.
Mazandaran Regional Water Company (MRWC), 2020. Hydrogeologic Studies, the Monthly Data of Pizeometric Wells. Mazandaran plain.
Mosaffa, M., Nazif, S., Amirhosseini, Y.K., Balderer, W., Meiman, M.H., 2021. An investigation of the source of salinity in groundwater using stable isotope tracers and GIS: a case study of the Urmia Lake basin, Iran. Groundwater Sustain. Dev. 12, 100513, 2021.
Motevalli, A., Naghibi, S.A., Hashemi, H., Berndtsson, R., Pradhan, B., Gholami, V., 2019. Inverse method using boosted regression tree and k-nearest neighbor to quantify effects of point and non-point source nitrate pollution in groundwater. J. Clean Prod. 228, 1248–1263.
Nemcic-Jurec, J., Jazbec, A., 2017. Point source pollution and variability of nitrate concentrations in water from shallow aquifers. Appl. Water Sci. 7, 1337–1348. https://doi.org/10.1007/s13201-015-0369-9.
Parvizishad, M., Dalvand, A., Mahvi, A.H., Goodarzi, F., 2017. A review of adverse effects and benefits of nitrate and nitrite in drinking water and food on human health. Health Scope 6 (3), e14164. https://doi.org/10.5812/jhealthscope.14164.
Qin, D., Qian, Y., Han, L., Wang, Z., Li, C., Zhao, Z., 2011. Assessing impact of irrigation water on groundwater recharge and quality in arid environment using CFCs, tritium and stable isotopes, in the Zhangye Basin, Northwest China. J. Hydrol. 405 (1–2), 194–208.
Rahmati, O., Tahmasebipour, N., Haghizadeh, A., Pourghasemi, H.R., Feizizadeh, B., 2017. Evaluation of different machine learning models for predicting and mapping the susceptibility of gully erosion. Geomorphology 298, 118–137. https://doi.org/10.1016/j.geomorph.2017.09.006.
Rahmati, O., Choubin, B., Fathabadi, A., Coulon, F., Soltani, E., Shahabi, H., Mollaefar, E., Tiefenbacher, J., Cipullo, S., Bin Ahmad, B., Tien Bui, D., 2019. Predicting uncertainty of machine learning models for modelling nitrate pollution of groundwater using quantile regression and UNEEC methods. Sci. Total Environ. 688, 855–866. https://doi.org/10.1016/j.scitotenv.2019.06.320.
Reiser, G., 2004. Evaluation of Streamflow, Water Quality, and Permitted and Nonpermitted Loads and Yields in the Raritan River Basin, vols. 1991–98. Water Years, New Jersey, p. 210.
Saemi, M., Ahmadi, M., 2008. Integration of genetic algorithm and a coactive neuro-fuzzy inference system for permeability prediction from well logs data. Transport Porous Media 71 (3), 273–288. https://doi.org/10.1007/s11242-007-9125-4.
Sajedi-Hosseini, F., Malekian, A., Choubin, B., Rahmati, O., Cipullo, S., Coulon, F., Pradhan, B., 2018. A novel machine learning-based approach for the risk assessment of nitrate groundwater contamination. Sci. Total Environ. 644 (10), 954–962. https://doi.org/10.1016/j.scitotenv.2018.07.054.
Savard, M.M., Paradis, D., Somers, G., Liao, S., Bochove, E.V., 2007. Winter nitrification contributes to excess NO3 in groundwater of an agricultural region: a dual-isotope study. Water Resour. Res. 43, W06422. https://doi.org/10.1029/2006WR005469.
Shi, X., Jiang, S., Xu, H., Jiang, F., He, Z., Wu, J., 2016. The effects of artificial recharge of groundwater on controlling land subsidence and its influence on groundwater quality and aquifer energy storage in Shanghai, China. Environ. Earth Sci. 75, 195. https://doi.org/10.1007/s12665-015-5019-x, 2016.
Shi, W., Lu, C., Werner, A.D., 2021. Assessment of the impact of sea-level rise on seawater intrusion in sloping confined coastal aquifers. J. Hydrol., 124872 https://doi.org/10.1016/j.jhydrol.2020.124872.
Shivasorupy, B., Barry, J., Mathias Maier, L., 2012. Sanitary hazards and microbial quality of open dug wells in the Maldives islands. J. Water Resour. Protect. (4), 474–486.
Taneja, P., Labhasetwar, P., Nagarnaik, P., Ensink, J.H., 2017. The risk of cancer as a result of elevated levels of nitrate in drinking water and vegetables in Central India. J. Water Health 15 (4), 602–614. https://doi.org/10.2166/wh.2017.283, 2017.
Tongal, H., Booij, M.J., 2018. Simulation and forecasting of streamflows using machine learning models coupled with base flow separation. J. Hydrol. 564, 266–282.
Tweed, S.O., Leblanc, M., Webb, J.A., Lubczynski, M.W., 2007. Remote sensing and GIS for mapping groundwater recharge and discharge areas in salinity prone catchments, southeastern Australia. Hydrogeol. J. 15, 75–96.
Valentini, M., dos Santos, G.B., Muller Vieira, B., 2021. Multiple linear regression analysis (MLR) applied for modeling a new WQI equation for monitoring the water quality of Mirim Lagoon, in the state of Rio Grande do Sul—Brazil. SN Appl. Sci. 3, 70. https://doi.org/10.1007/s42452-020-04005-1, 2021.
Wang, X., Zhang, F., Ding, J., 2017. Evaluation of water quality based on a machine learning algorithm and water quality index for the Ebinur Lake Watershed, China. Sci. Rep. 7, 12858. https://doi.org/10.1038/s41598-017-12853-y.
Wells, M.J., Gilmore, T.E., Nelson, N., Mittelstet, A., Böhlke, J.K., 2021. Determination of vadose zone and saturated zone nitrate lag times using long-term groundwater monitoring data and statistical machine learning. Hydrol. Earth Syst. Sci. 25, 811–829. https://doi.org/10.5194/hess-25-811-2021.