CrossMark

Comparison of machine learning models for predicting fluoride contamination in groundwater

Rahim Barzegar[1] · Asghar Asghari Moghaddam[1] · Jan Adamowski[2] ·
Elham Fijani[3]

Abstract Groundwater is an especially important freshwater source for water supplies in the Maku area of northwest Iran. The groundwater of the area contains high concentrations of fluoride and is, therefore, important in predicting the fluoride contamination of the groundwater for the purpose of planning and management. The present study aims to evaluate the ability of the extreme learning machine (ELM) model to predict the level of fluoride contamination in the groundwater in comparison to multilayer perceptron (MLP) and support vector machine (SVM) models. For this purpose, 143 water samples were collected in a five-year period, 2004–2008. The samples were measured and analyzed for electrical conductivity, pH, major chemical ions and fluoride. To develop the models, the data set—including Na^+, K^+, Ca^{2+} and HCO_3^- concentrations as the inputs and fluoride concentration as the output—was divided into two subsets; training/validation (80% of data) and testing (20% of data), based on a cross-validation technique. The radial basis-based ELM model resulted in an R^2 of 0.921, an NSC of 0.9071, an RMSE of 0.5638 (mg/L) and an MABE of 0.4635 (mg/L) for the testing data. The results showed that the ELM models performed better than MLP and SVM models for prediction of fluoride contamination. It was observed that ELM models learned faster than the other models during model development trials and the SVM models had the highest computation time.

✉ Rahim Barzegar
rm.barzegar@yahoo.com

1   Department of Earth Sciences,Faculty of Natural Sciences, University of Tabriz, Tabriz, Iran

2   Department of Bioresource Engineering, Faculty of Agricultural and Environmental Sciences, McGill University, Ste Anne de Bellevue, QC H9X 3V9, Canada

3   School of Geology, University College of Science, University of Tehran, Tehran, Iran

1 Introduction

Fluoride is an important constituent in groundwater because, among other things, it is required ..... for the healthy growth of bones and teeth in human beings (Asghari Moghaddam and Fijani 2008; Rafique et al. 2008). However, long-term intake of high doses of fluoride can have adverse effects on human health and result in fluorosis, a bone disorder (Cerklewski 1997; Barbier et al. 2010; Patel et al. 2014). The permissible limit for fluoride concentration in water is 1.5 mg/L according to the World Health Organization guidelines (WHO 2008). Groundwater contamination with fluoride is a serious worldwide problem that has negative effects on public health; globally, around 200 million people from 25 nations are exposed to grave health risks because of high amounts of fluoride in groundwater (Ayoob and Gupta 2006).

Fluoride occurs in almost all natural waters from trace concentrations to as high as 15,000 mg/L in mine water from the Kola Peninsula (Kraynov et al. 1969; Valenzuela-Vasquez et al. 2006). The natural concentration of fluoride in groundwater is governed principally by climate, composition of the host rock, and hydrogeology (Gupta et al. 2006; Valenzuela-Vasquez et al. 2006). High concentrations of fluoride in the groundwater are also contributed by anthropogenic activities such as the use of phosphatic fertilizers, pesticides, sewage and sludge, as well as

🖄 Springer

depletion of the groundwater table (EPA 1997; Ramanaiah et al. 2006; Kundu and Mandal 2009). Many factors can control the release of fluoride to groundwater, including the mineralogy of the rock (e.g. granite rocks), hydrogeological conditions, groundwater chemistry (e.g. presence or absence of ion complexes, precipitation of ions and colloids, and anion exchange capacity of aquifer materials), the interaction period of groundwater with a particular formation, and the dissolution kinetics of fluoride-bearing minerals (Patel et al. 2014).

Groundwater quality modeling enables the identification of groundwater quality trends and their influencing variables, which are important components of water resources management. In the last decades, numerically and physically based groundwater models were the most common groundwater modeling tools (Javadi and Al-Najjar 2007). However, the performance of these models depends on an adequate understanding of the hydrological behavior of the process in question, and the availability of detailed data on groundwater system properties. These two conditions are often absent, especially in developing regions, resulting in unsatisfactory model performance (Coppola et al. 2005; Alagha et al. 2014).

Numerical models are employed to simulate hydrological and hydrogeological problems, but these models are less user-friendly and lack knowledge transfer in model interpretation, which is leading to a large gap between model developers and practitioners. The advancement in Artificial Intelligence (AI) over the past two decades makes it possible to integrate these technologies into numerical modeling systems in order to bridge the gaps (Chau 2006). Also, AI techniques have rendered it possible to simulate human problem—solving expertise in this narrowly—defined domain by integrating descriptive knowledge, procedural knowledge and reasoning knowledge (Mirabbasi 2015; Chau 2006).

Recently, AI models have been used to predict groundwater contamination. For instance, Chowdhury et al. (2010) developed artificial neural network (ANN) models for spatial mapping of arsenic contamination of groundwater in Bangladesh. Alagha et al. (2014) applied AI models such as ANN and support vector machine (SVM) to predict nitrate contamination of the Gaza coastal aquifer. Cho et al. (2011) developed an ANN model for prediction of contamination potential of groundwater arsenic in Cambodia, Laos and Thailand. Al-Mahallawi et al. (2012) used neural networks for the prediction of nitrate groundwater contamination in rural and agricultural areas. Sahoo et al. (2006) applied ANN to assess pesticide contamination of shallow groundwater in Illinois, USA. Sirat (2013) applied backpropagation neural networks (BP-NN) to data taken from 1302 domestic and rural hydraulic wells in the Mid-continent of the USA, including Illinois, Iowa and 12

other states to predict contamination of groundwater with pesticides.

Some researchers have used AI models for fluoride contamination of groundwater. For example, Dar et al. (2012) applied ANNs for fluoride contamination of the Mamundiyar basin, India. Amini et al. (2009) used several hybrid methods by combining two classification techniques, classification tree and knowledge based clustering, and three predictive techniques (multiple regression, logistic regression and adaptive neuro-fuzzy inference system) for groundwater fluoride modeling using a global fluoride database. Nadiri et al. (2013) used a supervised committee machine artificial intelligence model for predicting groundwater fluoride concentrations of the Maku area. Chitsazan et al. (2016) applied hierarchical Bayesian model averaging to combine the predictions of multiple artificial neural networks (ANNs) for fluoride contamination of the Maku area. As can be seen, AI models are generally able to predict the contamination of groundwater. However, to date, no research has been published that uses an extreme learning machine (ELM) model to predict groundwater quality, especially groundwater contamination. For other applications, Zhang et al. (2015) proposed a self-adaptive differential evolution extreme learning machine (SADE-ELM) model for classification of water quality parameters in the Huaihe River, China. Imen (2015) applied artificial neural network, ELM and genetic programming for the long-term observation of total organic carbon (TOC) concentrations throughout Lake Mead in the United States. Dongwen (2013) used ELM to forecast total phosphorous and total nitrogen of a reservoir in Yunnan province, China.

Earlier studies in the Maku area (Asghari Moghaddam and Fijani 2008, 2009; Asghari Moghaddam et al. 2005, 2007) have indicated high concentrations of fluoride in the groundwater. The main objective of the present study is to investigate the ability of an extreme learning machine to predict the fluoride contamination of groundwater in the Maku area of northwest Iran. The usefulness of the ELM model was verified against the multilayer perceptron and support vector machine (SVM) models.

2 Methodology

2.1 Multilayer perceptron

A detailed description of ANN models is given in Haykin (Haykin 1999). However, in brief, ANNs consist of an input layer of source nodes, one or more hidden layers of computation nodes or neurons and one output layer. The input layer nodes distribute the input information to the next layer (i.e. the first hidden layer). The hidden and

output layer nodes process all incoming signals by applying factors to them (termed weights). Each layer also has an additional element called a bias node. Bias nodes simply output a signal to the nodes of the current layer. All inputs to a node are weighted, combined and then processed through a transfer function that controls the strength of the signal released through the node's output connections. Some of the most popular transfer (activation) functions are Sigmoid, Gaussian, Hyperbolic, Tangent and Hyperbolic Secant (Malekmohammadi et al. 2011; Barzegar et al. 2016c).

The proposed method for the ANN to be used in this study was the Multilayer Perceptron, in which the neurons are disposed in successive layers (feed-forward). Backpropagation is the most popular algorithm used for training a feed-forward ANN (Jain and Srinivasulu 2004; Fernando and Shamseldin 2009; Goyal et al. 2014). The structure of the MLP neural network model is shown in Fig. 1. In this figure, i, j and k denote input layer, hidden layer and output layer neurons, respectively, and w is the applied weight by the neuron. The explicit expression for an output value of a three-layered MLP is given by Belayneh and Adamowski (2012), Nourani et al. (2013), Barzegar and Asghari Moghaddam (2016) and Barzegar et al. (2016b, c):

$$y_k = f_\circ \left[\sum_{i=1}^{M_N} W_{kj} \cdot f_h \left(\sum_{i=1}^{N_N} W_{ji} X_i + W_{j0} \right) + W_{k0} \right] \tag{1}$$

where W_{ji} is a weight in the hidden layer connecting the ith neuron in the input layer and the jth neuron in the hidden layer, W_{j0} is the bias for the jth hidden neuron, f_h is the activation function of the hidden neuron, W_{kj} is a weight in the output layer connecting the jth neuron in the hidden layer and the kth neuron in the output layer, W_{k0} is the bias for the kth output neuron, f_o is the activation function for the output neuron, X_i is the ith input variable for the input layer and y_j is the computed output variable. N_N and M_N are the number of the neurons in the input and hidden layers,
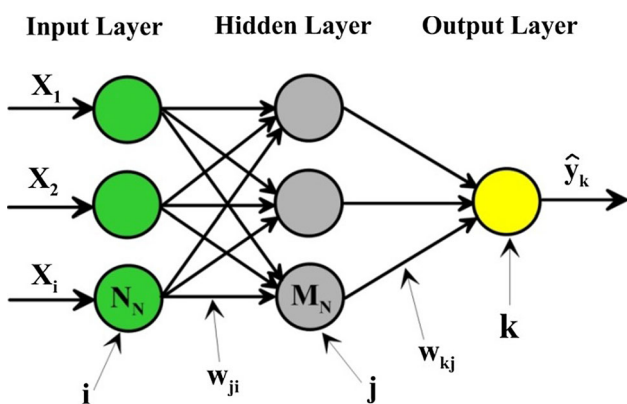
respectively. The gradient descent, conjugate gradient, Levenberg–Marquardt, and other learning algorithms can be used for training the MLP model (Kisi et al. 2015; Barzegar and Asghari Moghaddam 2016).

2.2 Support vector machine

The support vector machine (SVM) is a popular estimator introduced by Vapnik (1995). Based on Vapnik's theory, the SVM functions are offered by Eqs. (2–6), where $R = \{x_i, d_i\}_i^n$ is used for assuming a set of data points, the input space vector of the data sample is shown by x_i, and the desired value and data size are defined as d_i and n, respectively. The procedure of an SVM regression estimator (f) is written as (Zaji et al. 2016; Amirmojahedi et al. 2016; Mojumder et al. 2016; Ebtehaj et al. 2016; Al-Shammari et al. Al-Shammari et al. 2016; Shamshirband et al. 2016):

$$f(x) = w\varphi(x) + b \tag{2}$$

$$R_{SVMs}(C) = \frac{1}{2}||w^2|| + C\frac{1}{n}\sum_{i=1}^{n} L(x_i, d_i) \tag{3}$$

where $\varphi(x)$ is a high dimensional space feature that maps the input space vector x, w is a weight vector, b is a bias and $C\frac{1}{n}\sum_{i=1}^{n} L(x_i, d_i)$ represents the empirical error. The parameters w and b can be estimated with a regularized risk minimization function after introducing positive slack variables ξ_i and ξ_i^*, which represent upper and lower excess deviation, respectively.

$$\text{Minimize} \quad R_{SVMs}(w, \xi^*) = \frac{1}{2}||w^2|| + C\sum_{i=1}^{n}(\xi_i - \xi_i^*) \tag{4}$$

$$\text{Subject to} \begin{cases} d_i - w\varphi(x_i) + b_i \leq \varepsilon + \xi_i \\ w\varphi(x_i) + b_i - d_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, \quad i = 1, \ldots, l \end{cases} \tag{5}$$

where $\frac{1}{2}||w^2||$ is the regularization term, C is the error penalty factor used to regulate the difference between the regularization term and empirical error, ε is the loss function, which equates to the approximation accuracy of the training data point and l is the number of elements in the training data set.

Equation (2) can be resolved by proposing a Lagrange multiplier and optimality constraints, therefore obtaining a generic function given by Eq. (6):

$$f(x, \beta_i\beta_i^*) = \sum_{i=1}^{n}(\beta_i - \beta_i^*)K(x_i, x_j) + b \tag{6}$$

where $K(x_i, x_j)$ is recognized as the kernel function and it is equal to $K(x_i, x_j) = \varphi(x_i)\varphi(x_j)$. The latter term is an inner product of the two vectors, x_i and x_j, in the feature space

Fig. 1 The structure of the MLP model

$\varphi(x_i)$ and $\varphi(x_j)$, respectively. This inner product space is a vector space that has an additional structure termed as the inner product. This relates each pair of vectors with a scalar quantity known as the inner product of the vectors. The structure of the SVM model is shown in Fig. 2.

In this study, a radial basis function (RBF) $K(x_i, x_j) = exp(-\gamma||x_i - x_j||), \gamma > 0, \gamma = 1/(2\sigma^2)$, linear function $K(x_i, x_j) = x_i x_j$, polynomial basis function $K(x_i, x_j) = [(x_i x_j) + c]^d (c \geq 0$, d is the degree of the polynomial kernel), and sigmoid function $K(x_i, x_j) = tanh(\gamma x_i x_j + c)(\gamma > 0, c \geq 0)$ were applied as the kernel functions.

2.3 Extreme learning machine

Extreme learning machine (ELM) was first proposed by Huang et al. (2004) as a fast learning technique with high generalization performance that uses single-hidden layer (feature mapping) feed-forward neural networks (SLFNs) (Huang et al. 2004, 2006; Abdullah et al. 2015). The ELM chooses the input weights randomly and determines the output weights of the SLFN analytically (Aghbashlo et al. 2016). It is capable of determining all the network parameters analytically, which prevents trivial human intervention (Shamshirband et al. 2016). The main objectives of the ELM are to reach the smallest training errors, the smallest norm of output weights, and good generalization performance (Huang et al. 2006).

The network structure of the ELM model is shown in Fig. 3. For N different training samples $(x_i, y_i) \in R^n \times R^m (i = 1, 2, 3, \ldots, n)$, the number of hidden nodes is L. The SLFN model, which has the activation function $f(x)$, can be expressed as (Ding et al. 2016; Liu et al. 2016; Aghbashlo et al. 2016):

$$\sum_{i=1}^{L} \beta_i f_i(x_j) = \sum_{i=1}^{\tilde{L}} \beta_i f(a_i \cdot b_i \cdot x), \quad j = 1, \ldots, N \quad (7)$$
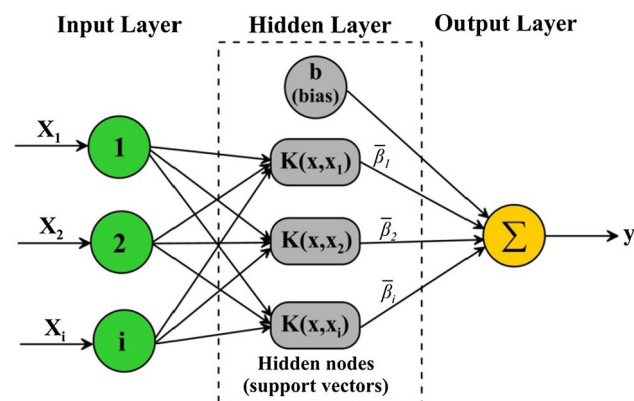
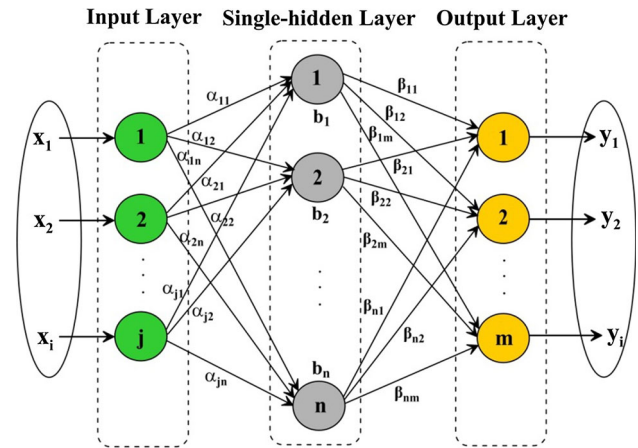**Fig. 2** The structure of the SVM model

Fig. 3 The structure of the ELM model

where $a_i = [a_{i1}, a_{i2}, \ldots, a_{in}]^T$ is the input weight vector connected to the hidden layer node, i, b_i is the bias value of hidden layer nodes, $\beta_i = [\beta_{i1}, \beta_{i2}, \ldots, \beta_{im}]^T$ are the output weight vectors connected to the hidden layer node, and i, a_i, x_j is the inner product of $a_i \cdot x_j$.

Equation (7) can be rewritten compactly as follows:

$$\sum_{i=1}^{L} \beta_i f_i(x_j) = H\beta \quad (8)$$

$$H = \begin{bmatrix} f(a_1 \cdot x_1 \cdot b_1) & \ldots & f(a_L \cdot x_1 \cdot b_L) \\ \vdots & \ldots & \vdots \\ f(a_1 \cdot x_N \cdot b_1) & \ldots & f(a_L \cdot x_N \cdot b_L) \end{bmatrix}_{N \times L} \quad (9)$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L \times m} \quad T = \begin{bmatrix} y_1^T \\ \vdots \\ y_L^T \end{bmatrix}_{N \times m} \quad (10)$$

where H is the output matrix of the hidden layer, β is the output weight matrix, and T is the label matrix.

Not all parameters need to be adjusted when the excitation function $f(x)$ is infinitely differentiable at any interval. At the start of the training process, SLFNs are assigned random values to the input weight a and hidden layer node bias b. When input weights and hidden layer node biases are determined by random assignment methods, the hidden layer output matrix H from the input samples can be obtained. Thus, training SLFNs are transformed into solving least square solutions.

By introducing regularization theory into the ELM model, the cost function can be expressed as:

$$minL_{ELM} = \frac{1}{2}||\beta||^2 + \frac{C}{2}||T - H\beta||^2 \quad (11)$$

The least squares solution of Eq. (11) is:

$$V - CH^T(T - H\beta) = 0 \quad (12)$$

When the number of training samples is more than the number of hidden layer nodes,

$$\beta = \left(\frac{1}{C} + H^T H\right)^{-1} H^T T \tag{13}$$

When the number of training samples is less than the number of hidden layer nodes,

$$\beta = H^T \left(\frac{1}{C} + H H^T\right)^{-1} T \tag{14}$$

When the number of hidden layer units is large enough in the ELM algorithm, the regression accuracy of the algorithm is always stable.

In this study, the activation functions were defined by sine (f_{sin}), sig (f_{sig}), hard-limit ($f_{hardlim}$), radial basis (f_{radbas}) and triangular basis (f_{tribas}), as in the following equations:

$$f_{sin} = (a, b, x) = \sin(ax + b) \tag{15}$$

$$f_{sig} = (a, b, x) = \frac{1}{1 + \exp(-(ax + b))} \tag{16}$$

$$f_{hardlim}(a, b, x) = \begin{cases} 1, & if\ ax + b < 0 \\ 0, & otherwise \end{cases} \tag{17}$$

$$f_{radbas}(a, b, x) = exp\left(-(ax + b)^2\right) \tag{18}$$

$$f_{tribas}(a, b, x) = \begin{cases} 1 - |ax + b|, & if\ -1 \le (ax + b) \le 1 \\ 0, & otherwise \end{cases} \tag{19}$$

2.4 Performance evaluation of the models

The performance of the developed models for training and testing sets was evaluated by following measures of goodness-of-fit: the coefficient of determination (R^2), Nash–Sutcliffe efficiency coefficient (NSC), root mean squared error (RMSE) and mean absolute bias error (MABE), shown in Eqs. (20–23), respectively. R^2 expresses the degree of the relation when two variables are linearly related. If R^2 is close to 1, there is good correlation between the observed and predicted values. The Nash–Sutcliffe coefficient of efficiency (NSC), an indicator of the model fit, is a normalized measure ($-\infty$ to 1) that compares the mean square error generated by a particular model simulation to the variance of the target output sequence. An NSC value of 1 indicates perfect model performance, an NSC value of zero indicates that the model is, on average, performing only as well as the use of the mean target value as prediction, and an NSC < 0 indicates an altogether questionable choice of the model (Nash and Sutcliffe 1970). A perfect fit between observed and predicted values would have an RMSE of 0.

$$R^2 = \left[\sum_{i=1}^{N}(P_i - \bar{P})(O_i - \bar{O})\right]^2 \left[\sum_{i=1}^{N}(P_i - \bar{P})^2 (O_i - \bar{O})^2\right]^{-1} \tag{20}$$

$$NSC = 1 - \frac{\sum_{i=1}^{N}(O_i - P_i)^2}{\sum_{i=1}^{N}(O_i - \bar{O})^2} \tag{21}$$

$$RMSE = \left[N^{-1} \sum_{i=1}^{N}(P_i - O_i)^2\right]^{0.5} \tag{22}$$

$$MABE = \frac{1}{N} \sum_{i=1}^{N}|P_i - O_i| \tag{23}$$

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

3 Study area and data

3.1 Study area

The Maku area is located in the north of West Azerbaijan province in the northwest of Iran. It lies between 44°21′ and 45°10′ east longitude and 35°10′ and 39°34′ north latitude, covering an area of approximately 1600 km². The study area is covered up to 400 km² by basaltic lavas. It is bounded in the west by Turkey and in the east by the Aras River, as shown in Fig. 4. The most important cities in the area are Maku, Poldasht and Bazargan. The climate of the area is cold and arid. The annual average precipitation is about 300 mm and the maximum and minimum precipitation occur in May and September, respectively (Asghari Moghaddam and Fijani 2009). Mean daily temperatures at the Maku Synoptic Station (1411 m amsl) vary from −7.4 °C in January up to 17.2 °C in July, with an annual average of 10.4 °C (Asghari Moghaddam and Fijani 2009). The main rivers in the study area are Sari Su and Zangmar, which flow from west to east.

The Maku area includes formations of Precambrian to Quaternary ages. The major formation in the area is lava, which mainly consists of basaltic rocks. The great extent of young lava in the extreme northwest of Iran is attributable to the volcanic activity of Ararat in Turkey (Asghari Moghaddam and Fijani 2008). Young alluvium clay to gravel sheets, spreading as fan deposits from the mountain flanks and flood plains, are the recent unconsolidated materials filling the lowlands and river beds (Asghari Moghaddam and Fijani 2007, 2008). The Maku area aquifers have a range of lithologies, including basaltic-alluvium, alluvium and karstified limestone. However, the basaltic-alluvium aquifer forms the main water-bearing
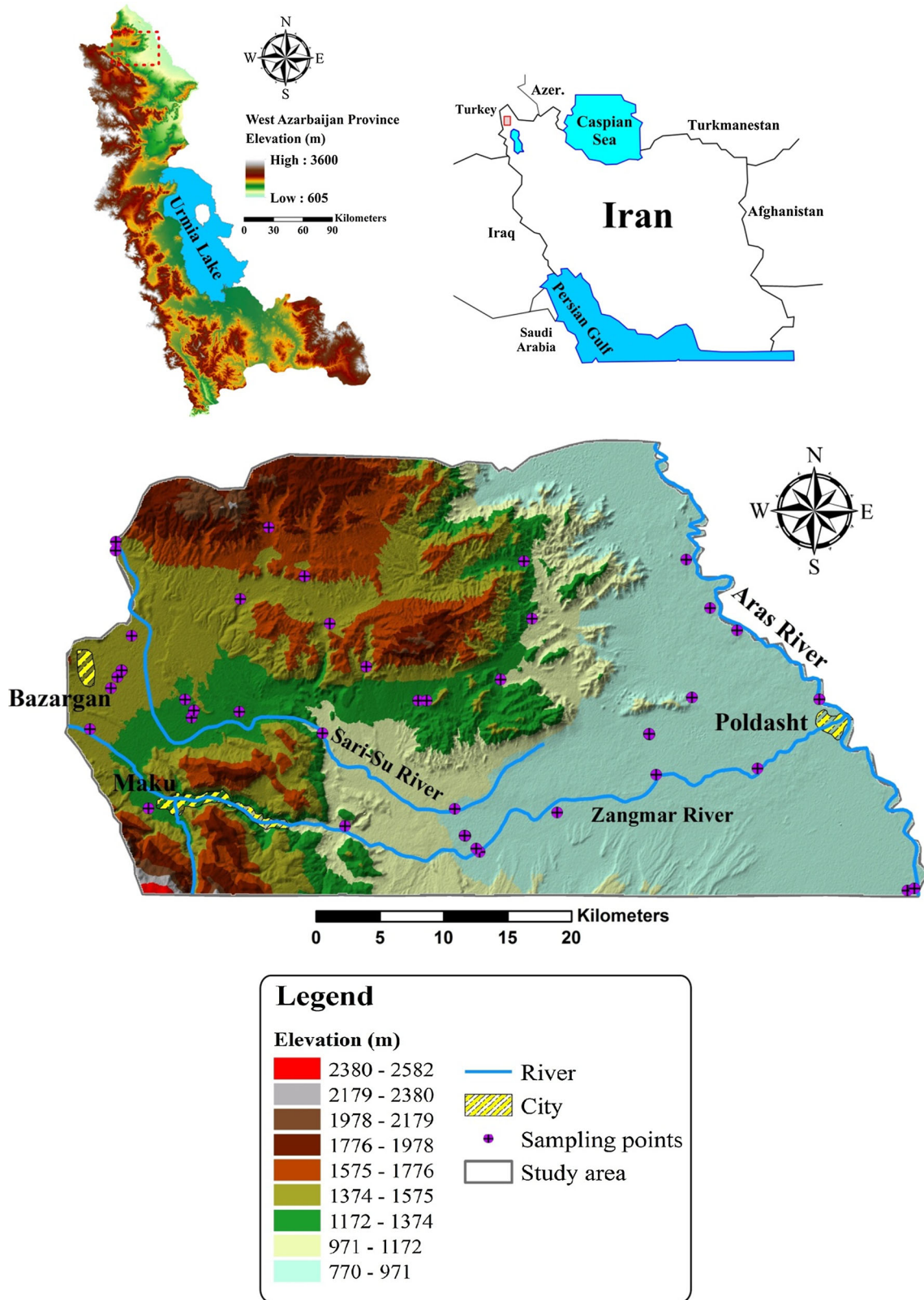
Fig. 4 Location of the study area and sampling points

layers of the area (Asghari Moghaddam and Fijani 2009). Detailed discussion regarding the geology, hydrogeology and hydrochemistry of the Maku area is presented in Asghari Moghaddam et al. (2005), Fijani (2007) and Asghari Moghaddam and Fijani (2007, 2008, 2009).

Groundwater is the main water source used for various purposes such as drinking, agriculture and industry in the study area. Fluoride contamination is considered to be the main water quality problem in the Maku area, where the average concentration of fluoride is 2.85 mg/L (Asghari Moghaddam and Fijani 2008, 2009). The volcanic rocks in the study area contain silicate minerals, apatite and fluorapatite, and the weathering of these minerals is likely to be the main source of fluoride in the groundwater of the study area (Asghari Moghaddam et al. 2005; Fijani 2007).

3.2 Data collection and pre-processing

In this study, the chemical analyses of 143 water samples were used. Water was collected from 39 sampling sites over water sources (wells, springs, qanats, etc.) in a 5-year period, from 2004 to 2008. The locations of the sampling sites are shown in Fig. 4. The largest and smallest data sets were in August 2006 (38 samples) and July 2004 (8 samples), respectively. The water samples were analyzed in the Hydrogeology Laboratory of the University of Tabriz. The electrical conductivity (EC) and pH were measured in situ in the field. Fluoride concentration in water samples was determined using the method of SPADNS—using a Spectro 40 spectrophotometer at 570 nm—and the other ions (Ca^{2+}, Mg^{2+}, Na^+, K^+, HCO_3^-, CO_3^{2-}, SO_4^{2-} and Cl^-) were determined by standard methods (American Public Health Association 1998). The accuracy of the water analysis was within the limit of $\pm 5\%$ according to the cation–anion balance (Domenico and Schwartz 1990). In this study complete hydrological data sets (e.g. groundwater flow and stream flow) were not available, except groundwater level, and there was no correlation between groundwater level and fluoride contamination. Therefore, this study may indicate the suitability of certain AI models for hydrological modeling, particularly in regions where detailed and complete data sets about hydrological processes are usually unavailable. For example, in this case study, there are many data sets of major ions, but the fluoride concentrations are not available for such data sets. Therefore, AI models can be used for prediction of the unavailable fluoride concentrations.

One of the most important steps in developing a prediction model is the selection of the input variables. For the selection of input variables, certain fluoride-related variables were chosen. For this purpose, principal component analysis (PCA) was used. PCA can be used to reduce the complexity of input variables when there are large volumes of information and better interpretation of variables is recommended (Noori et al. 2010, 2011). It can be seen from Table 1 that Na^+, K^+, Ca^{2+} and HCO_3^- concentrations have the greatest effect on the first component (PC1). Therefore, these four variables were selected as inputs of the developed models. The high positive loading of the HCO_3^- in PC1 could be due to the release of hydroxyl and bicarbonate ions simultaneously during the leaching and dissolution process of fluoride bearing minerals into the groundwater. Groundwater with high K^+ and Na^+ concentrations likely occurs as a result of silicate mineral hydrolysis of volcanic rocks as a source of fluoride in the study area. High concentrations of Na^+ increase the solubility of the fluoride bearing minerals. Also, the presence of Na^+, K^+ and HCO_3^- variables in PC1 could be an indication of ion-exchange and carbonate weathering. The negative loading of the Ca^{2+} reflects precipitation of fluorite (CaF_2) mineral, due to the high solubility product of fluoride (Rafique et al. 2008).

Before developing the models, the data set values were normalized between 0.2 and 0.8, using Eq. (24):

$$Normalized\, X = C_1 \frac{X_i - X_{min}}{X_{max} - X_{min}} + C_2 \qquad (24)$$

in which X_{max} and X_{min} are the maximum and minimum of the data sets. In the current study, the C_1 and C_2 values were assigned as 0.6 and 0.2, respectively. Therefore, the data were normalized into the range [0.2, 0.8]. This normalization was employed following the suggestion of Cigizoglu (2003), who showed that scaling input data between 0.2 and 0.8 gives ANNs the flexibility to estimate beyond the training range.

To develop the MLP, SVM and ELM models, the cross-validation technique (Chang et al. 2013; Fijani et al. 2013; Barzegar et al. 2016b) was used to divide the data sets into training and testing sub sets. The data sets were divided into a

Table 1 Principal component analysis (PCA) to determine the input variables

Variables	PC1	PC2	PC3
EC	0.547	0.720	−0.025
pH	0.221	0.045	0.826
Ca^{2+}	−0.767	0.160	−0.389
Mg^{2+}	0.104	0.861	−0.052
Na^+	0.877	0.370	0.066
K^+	0.841	−0.037	0.001
HCO_3^-	0.900	0.101	−0.174
CO_3^{2-}	0.258	−0.052	0.430
Cl^-	0.038	0.926	0.028
SO_4^{2-}	−0.201	0.707	−0.084
F^-	0.920	−0.087	0.080

training/validation set (80% of the data) and a testing set (the remaining 20% of the data). Statistical analysis of the training/validation and testing data sets are listed in Table 2.

4 Development of the models

4.1 MLP model

As previously mentioned, 80% of the data set was considered for training/validation and the remaining 20% for testing. For the MLP model, the training/validation set was further divided into 80% training and 20% validation, so overall, 64% of the data were used for training, 16% were used for validation, and 20% were used for testing.

The neural network training was implemented through the MATLAB Neural Network toolbox. In this study, the three-layered feed-forward neural network was trained with the Levenberg–Marquardt algorithm (TrainLM). This algorithm is a variation of Newton's method and is designed to second-order training speed without having to compute the Hessian matrix (Adamowski and Sun 2010). Traditionally, the trial and error method is used to select the optimal number of hidden neurons (Belayneh et al. 2014, 2016; Adamowski and Sun 2010; Barzegar et al. 2016a, b). However, Wanas et al. (1998) and Mishra and Desai (2006) empirically considered equations log(N), where N is the number of training samples, and $2n + 1$, where n is the number of input neurons to determine the number of hidden neurons. In this study, the optimal number of hidden neurons was determined to be between log (N) and $(2n + 1)$. Two hidden neurons and nine hidden neurons were determined by using the Wanas et al. (1998) and Mishra and Desai (2006) methods, respectively; thereafter, the optimal number was chosen via trial and error. The number of neurons in the hidden layer was seven. The transfer function between layer one and layer two was TANSIG, while PURELIN was used for the last layer. Learning rates and momentum factors of 0.1 and 0.2, respectively, were chosen by trial and error. The magnitude of the gradient and the number of validation checks used to terminate network training are presented in Fig. 5a. At an epoch of 16 iterations, the gradient was 7.069×10^{-4},

barely above the 1×10^{-4} threshold below which training will stop, and at six, the validation checks also indicated training should stop. The performance plot (Fig. 5b) shows the value of the function in terms of training, validation, and testing behaviors, versus the iteration number. The best validation performance was at epoch 10, based on a mean square error equal to 1.743×10^{-3}. The MLP model was trained in 0.45 s. When the training of the model was completed, the testing data served as model input and fluoride concentration values were predicted.

4.2 SVM model

In this study, DTREG (Data Regression) was utilized for the SVM modeling. The models were created by using the Epsilon-SVR kernel type. Both grid and pattern search—as well as tenfold cross-validation re-sampling methods—were employed to find optimal parameter values. During grid search, the program (DTREG) evaluates values of each parameter within the predefined search area. On the other hand, a pattern search (also known as a line search or compass search) starts from the centre of the search area and tries steps in both directions for each parameter. The centre of the search area is then moved to the new point if a better model fit is obtained. The process is repeated until the specified tolerance rate is reached (Sonebi et al. 2016; Al-Anazi and Gates 2010).

Model parameters such as C have a search range of 0.1–5000, kernel parameter γ of 0.001–50, and ε (Epsilon) of 0.001–100. By selecting the pattern search technique using 10 search intervals which would require 1000 model evaluations and 1e-008 tolerance for stopping the iterative optimization process and the optimal values through the grid search, we could create a fluoride prediction model with higher stability and lower RMSE. The optimal calibration constants and kernel parameters for developing SVM models are shown in Table 3. After constructing the SVM models, the testing data set was used for testing the models.

4.3 ELM model

The ELM models were developed in a MATLAB environment. Three layers were used to build the architecture

Table 2 Statistical analysis of the training/validation and testing data sets

Variables	Training/validation data set				Testing data set			
	Min	Mean	Max	SD	Min	Mean	Max	SD
Na^+	0.7	138.2	325.7	87.4	1.4	145.1	335.6	101.6
K^+	0.7	8.6	18.5	4.4	1.3	8.45	17.9	5.2
Ca^{2+}	5.6	73.3	146.8	2.8	19.2	67.61	123.5	26.6
HCO_3^-	187.7	565.8	914.4	2.1	104.8	552.1	857.2	224.1
F^-	0.2	2.7	7.9	1.7	0.46	2.9	6.4	1.8

Fig. 5 Training state and performance of the developed MLP neural network model
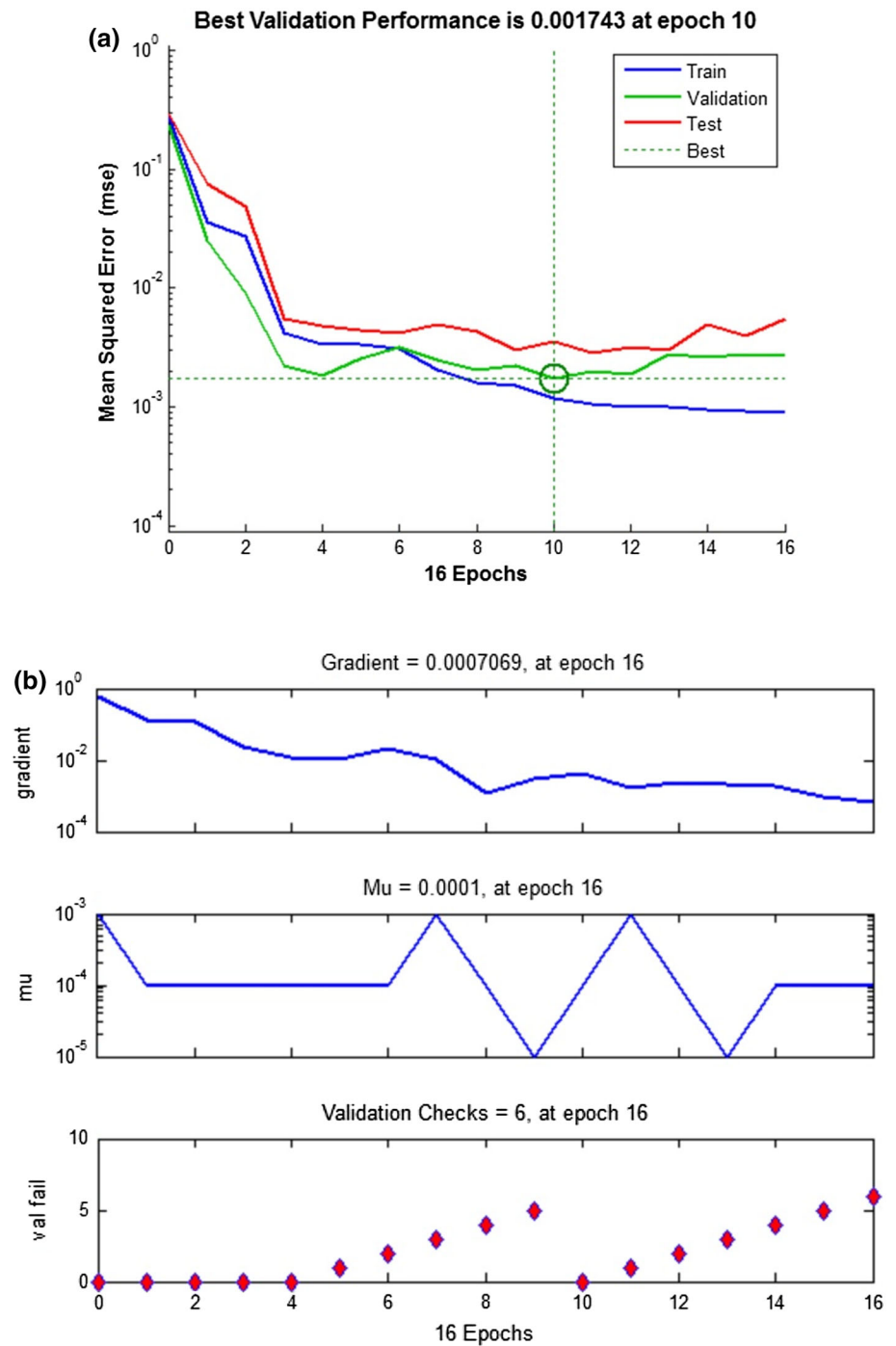
Table 3 Optimal training constants and kernel parameters for developing SVM models

Kernel function	Calibration constants		Kernel parameters			Time (s)
	ε	C	C	γ	d	
Linear	0.001	3865.48	–	–	–	12.32
Sigmoid	0.001	449.68	0.07	0.003	–	13.33
Polynomial	0.001	0.1	0.06	4.516	3	76.94
RBF	0.001	0.178	–	14.674	–	24.56

Table 4 Optimal hidden neurons for developing ELM models

Activation function	Hidden neurons	Time (s)
Sigmoid	20	0.015
Sine	22	0.01
Hard-limit	42	0.124
Radial basis	16	0.124
Triangular basis	16	0.062

for fluoride contamination prediction in designing the ELM. The number of neurons was 4 (input) for each of the developed ELM models. The ELM output layer had one neuron representing the predicted fluoride. The number of hidden nodes is changeable for better accuracy, and the number of hidden neurons was selected via a trial and error method. The number of neurons between 1 and 50 were tested in hidden layers of the ELM models. In each trial, the number of nodes in the hidden layer was increased gradually until the optimal nodes were reached. A taxonomy of activation functions was tried one-by-one, which included "sigmoid", "sine", "radial basis", "triangular basis" and "hard-limit". The optimal hidden neurons for different activation functions are listed in Table 4. After

training the models, the testing data set was used to test the developed models.

5 Results

The performance of the MLP, SVM and ELM models for prediction of fluoride contamination in both training and testing stages is presented in Tables 5, 6 and 7, respectively. The statistical evaluation criteria revealed that all the models for the prediction of fluoride concentration yielded satisfactory results. Therefore, these models are acceptable—due to high R^2 and NSC values and low RMSE values—for predicting fluoride contamination in the Maku area. The values of R^2 and NSC which were close to unity, and fairly low RMSE and MABE in all the models—for both the training and testing sets—emphasized good generalization and predictive abilities of the three modeling approaches for the given data set. However, relatively lower prediction errors obtained by models in the training set as compared to the testing set indicated that these models exhibited relatively better generalization as compared to the predictions.

**Table 5** Results of MLP model during training and testing

Model	Training				Testing			
	R^2	NSC	RMSE (mg/L)	MABE (mg/L)	R^2	NSC	RMSE (mg/L)	MABE (mg/L)
MLP	0.9191	0.9179	0.4914	0.3607	0.8152	0.8019	0.8232	0.6131

Table 6 Results of SVM models for training and testing stages

Kernel function	Training				Testing			
	R^2	NSC	RMSE (mg/L)	MABE (mg/L)	R^2	NSC	RMSE (mg/L)	MABE (mg/L)
Linear	0.8721	0.8885	0.5925	0.4316	0.8521	0.7774	0.8727	0.7124
Sigmoid	0.8851	0.8806	0. 5727	0. 4097	0.8524	0.7793	0.8688	0.7078
Polynomial	0.8932	0.891	0.5661	0.4001	0.8709	0.7866	0.8543	0.7181
RBF	0.9014	0.9122	0.5082	0.3146	0.8833	0.8658	0.6775	0.5596

Table 7 Results of ELM models for training and testing stages

Activation function	Training				Testing			
	R^2	NSC	RMSE (mg/L)	MABE (mg/L)	R^2	NSC	RMSE (mg/L)	MABE (mg/L)
Sigmoid	0.9213	0.92	0.4852	0.3812	0.901	0.8899	0.6138	0.522
Sine	0.9229	0.9213	0.4812	0.3789	0.9084	0.8947	0.6002	0.4988
Hard-limit	0.913	0.9126	0.5069	0.3907	0.901	0.8863	0.6236	0.4925
Radial basis	0.9453	0.9449	0.4024	0.3157	0.921	0.9071	0.5638	0.4635
Triangular basis	0.926	0.9254	0.4684	0.3584	0.9087	0.9018	0.5797	0.4772
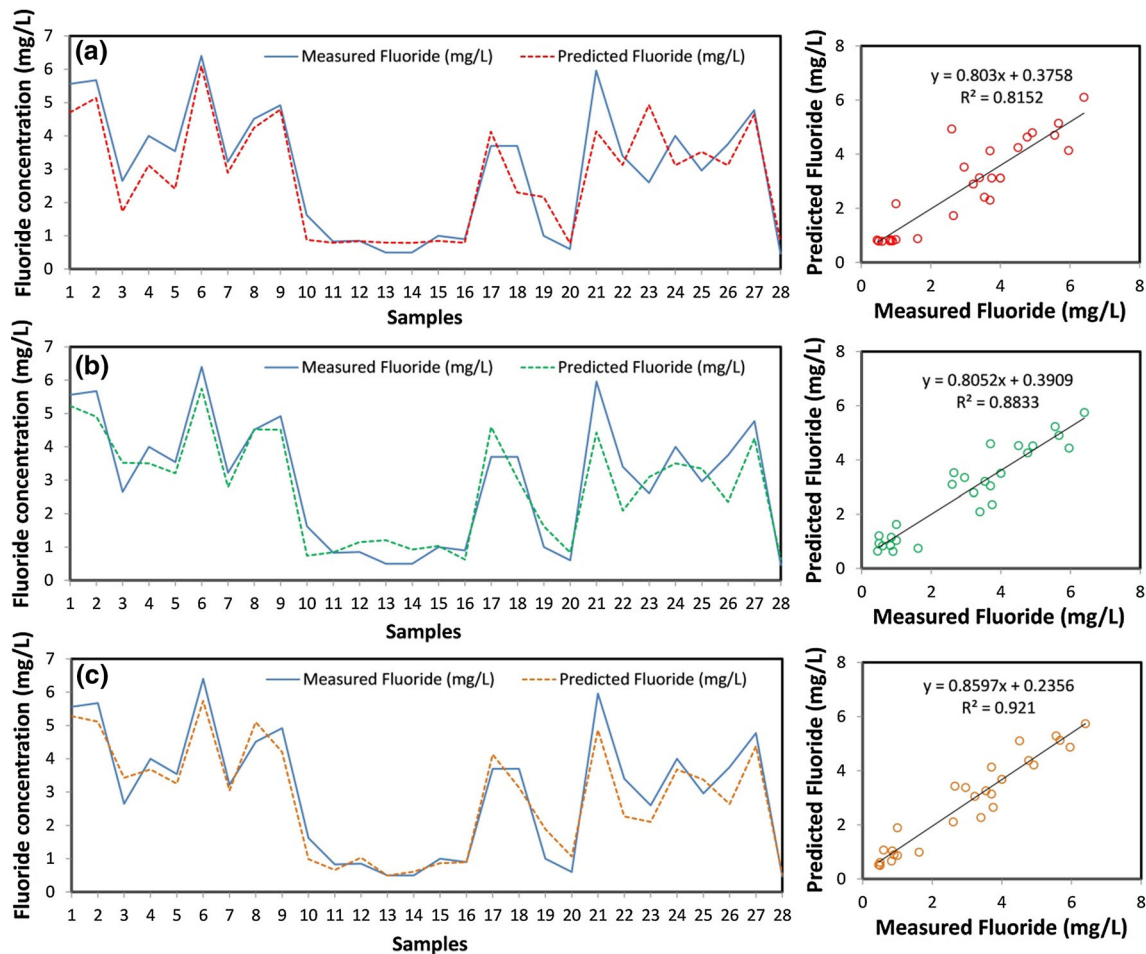
Fig. 6 The performance of the **a** MLP, **b** SVM and **c** ELM models developed for prediction of fluoride concentration in the testing stage

Table 5 shows the statistical evaluation criteria of training and testing for the MLP model. The R^2, NSC, RMSE and MABE of the MLP model for the training data were 0.9191, 0.9179, 0.4914 (mg/L) and 0.3607 (mg/L), respectively; those for the testing data were 0.8152, 0.8019, 0.8232 (mg/L) and 0.6131 (mg/L), respectively. Figure 6a shows the comparison between the measured and predicted values of fluoride for the MLP model in the testing stage.

Table 6 shows a performance comparison of the different kernel functions used for SVM model development. The RBF and linear kernel functions showed the best and worst performance, respectively, among the utilized kernel functions for the SVM models. The R^2, NSC, RMSE and MABE for the linear-based SVM model for training data were 0.8721, 0.8885, 0.5925 (mg/L) and 0.4316 (mg/L), respectively, whereas those for the testing data were 0.8521, 0.7774, 0.8727 (mg/L) and 0.7124 (mg/L), respectively. In the training stage, the SVM model with the RBF kernel function resulted in an R^2 of 0.9014, an NSC of 0.9122, an RMSE of 0.5082 (mg/L), and an MABE of

0.3146 (mg/L). However, for the testing data, the corresponding values were 0.8833, 0.8658, 0.6775 (mg/L) and 0.5596 (mg/L), respectively. The RBF kernel function significantly reduced the overall prediction errors. It was demonstrated that the radial basis kernel function (RBF) performed better than linear, sigmoid and polynomial kernel functions in terms of performance criteria. This result was confirmed by Rajasekaran et al. (2008), Yang et al. (2009), Wu and Wang (2009) and Amirmojahedi et al. (Amirmojahedi et al. 2016). Figure 6 shows a comparison between the measured and predicted values of the fluoride concentration for the SVM model with the RBF kernel function in the testing stage. The results show that the use of nonlinear kernel functions achieved better performance than the linear kernel.

Table 7 shows a comparison of the performance of the different activation functions utilized for ELM model development. The radial basis and hard-limit functions showed the best and worst performance among the activation functions used for ELM models, respectively. The R^2, NSC, RMSE and MABE for the hard-limit-based ELM

model for training data were 0.913, 0.9126, 0.5069 and 0.3907 (mg/L), respectively, whereas those for the testing data were 0.901, 0.8863, 0.6236 and 0.4925 (mg/L), respectively. The radial basis-based ELM model resulted in an R^2 of 0.9453, NSC of 0.9449, RMSE of 0.4024 (mg/L), and MABE of 0.3157 (mg/L) for the training data and in an R^2 of 0.921, NSC of 0.9071, RMSE of 0.5638 (mg/L) and MABE of 0.4635 (mg/L) for the testing data. The performance of the ELM with the radial basis function for fluoride contamination prediction in the testing stage is shown in Fig. 6c. The results show that the ELM models performed better than MLP and SVM models for prediction of fluoride contamination. Empirical studies have shown that the generalization ability of ELM is better than that of SVM models (Huang et al. 2006, 2012; Fernandez-Delgado et al. 2014; Huang et al. 2014, 2015).

The ELM models had advantages in computation time in comparison with MLP and SVM models. It was observed that ELM models learned faster than the other models during model development trials while the SVM models had the highest computation time. As analyzed by Huang et al. (2015), the training of SVM is a quadratic programming problem, and thus, it requires high computational costs. In contrast, the parameters of the ELM hidden layer need not be adjusted and can be independent of the training data. Hence, the ELM model only computes the output weights analytically, and it has a much faster learning speed and lower computational complexity than SVM (Wang et al. 2015). The grid search was another reason for the high computation times of the SVM models because, by using this method, the model must be evaluated at many points within the grid for each parameter (Al-Anazi and Gates 2010).

6 Conclusions

This study investigated the ability of three different machine learning algorithms including MLP, SVM and ELM to predict the fluoride contamination of groundwater in the Maku area of northwest Iran. The results demonstrated that the ELM models outperformed the MLP and SVM models for prediction of fluoride contamination. This study found that the SVM model with the RBF kernel function outperformed the linear-, sigmoid- and polynomial kernel function-based models. The radial basis and hard-limit functions, among the utilized activation functions, showed the best and worst performance for ELM models, respectively. During model development trials, it was observed that ELM models learned faster than the other models while the SVM models had the highest computation time.

References

Abdullah SS, Malek MA, Abdullah NS, Kisi O, Yap KS (2015) Extreme learning machines: a new approach for prediction of reference evapotranspiration. J Hydrol 527:184–195

Adamowski JF, Sun K (2010) Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. J Hydrol 390(1–2):85–91

Aghbashlo M, Shamshirband S, Tabatabaei M, Yee PL, Larimi YN (2016) The use of ELM–WT (extreme learning machine with wavelet transform algorithm) to predict exergetic performance of a DI diesel engine running on diesel/biodiesel blends containing polymer waste. Energy 94:443–456

Alagha JS, Said MAM, Mogheir Y (2014) Modeling of nitrate concentration in groundwater using artificial intelligence approach—a case study of Gaza coastal aquifer. Environ Monit Assess 186(1):35–45

Al-Anazi AF, Gates ID (2010) Support vector regression for porosity prediction in a heterogeneous reservoir: a comparative study. Comput Geosci 36:1494–1503

Al-Mahallawi K, Mania J, Hani A, Shahrour I (2012) Using of neural networks for the prediction of nitrate groundwater contamination in rural and agricultural areas. Environ Earth Sci 65(3):917–928

Al-Shammari ET, Keivani A, Shamshirband S, Mostafaeipour A, Yee PL, Petkovic D, Sudheer CH (2016) Prediction of heat load in district heating systems by support vector machine with Firefly searching algorithm. Energy 95:266–273

American Public Health Association (1998) Standard method for the examination of water and wastewater, 17th edn. American Public Health Association, Washington

Amirmojahedi A, Mohammadi K, Shamshirband S, Danesh AS, Mostafaeipour A, Kamsin A (2016) A hybrid computational intelligence method for predicting dew point temperature. Environ Earth Sci 75:415. doi:10.1007/s12665-015-5135-7

Amini M, Johnson A, Abbaspour KC, Mueller K (2009) Modeling large scale geogenic contamination of groundwater, combination of geochemical expertise and statistical techniques. 18th World IMACS/MODSIM Congress, Cairns, pp 4100–4106

Asghari Moghaddam A, Fijani E (2007) Hydrogeologic and hydrochemical studies of Maku Area basaltic and karstic aquifers in relation to geological formations (in Persian). Sci Q J Geosci 17(67):2–13

Asghari Moghaddam A, Fijani E (2008) Distribution of fluoride in groundwater of Maku area, northwest of Iran. Environ Geol 56:281–287

Asghari Moghaddam A, Fijani E (2009) Hydrogeologic framework of the Maku area basalts, northwestern Iran. Hydrogeol J 17:949–959

Asghari Moghaddam A, Jomeiri R, Mohamadi A (2005) Investigation of hydrogeologic characteristics of Maku area basaltic rocks for optimum groundwater management using geophysical, RS and GIS methods (in Persian). University of Tabriz, Tabriz

Asghari Moghaddam A, Jomeiri R, Mohamadi A (2007) Source of high fluoride in groundwater of basaltic lavas of Bazargan–Poldasht Plains and its ill effects on human health (in Persian). J Environ Stud (University of Tehran) 33:25–32

Ayoob S, Gupta AK (2006) Fluoride in drinking water: A review on the status and stress effects. Crit Rev Environ Sci Technol 36:433–487

Barbier O, Arreola-Mendoza L, Del Razo LM (2010) Molecular mechanisms of fluoride toxicity. Chem Biol Interact 188:319–333

Barzegar R, Adamowski J, Asghari Moghaddam A (2016a) Application of wavelet-artificial intelligence hybrid models for water

quality prediction: a case study in Aji-Chay River, Iran. Stoch Environ Res Risk Assess. doi:10.1007/s00477-016-1213-y

Barzegar R, Asghari Moghaddam A (2016) Combining the advantages of neural networks using the concept of committee machine in the groundwater salinity prediction. Model Earth Syst Environ. doi:10.1007/s40808-015-0072-8

Barzegar R, Asghari Moghaddam A, Baghban H (2016b) A supervised committee machine artificial intelligent for improving DRASTIC method to assess groundwater contamination risk: a case study from Tabriz plain aquifer, Iran. Stoch Environ Res Risk Assess 30(3):883–899

Barzegar R, Sattarpour M, Nikudel MR, Asghari Moghaddam A (2016c) Comparative evaluation of artificial intelligence models for prediction of uniaxial compressive strength of travertine rocks, case study: Azarshahr area, NW Iran. Model Earth Syst Environ. doi:10.1007/s40808-016-0132-8

Belayneh A, Adamowski J (2012) Standard Precipitation Index drought forecasting using neural networks, wavelet neural networks, and support vector regression. Appl Comput Intell Soft Comput 2012:794061. doi:10.1155/2012/794061

Belayneh A, Adamowski J, Khalil B (2016) Short-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet transforms and machine learning methods. Sustain Water Resour Manag 2(1):87–101

Belayneh A, Adamowski J, Khalil B, Ozga-Zielinski B (2014) Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet-support vector regression models. J Hydrol 508:418–429

Cerklewski FL (1997) Fluoride bioavailability: nutritional and clinical aspects. Nutr Res 17:907–929

Chang FJ, Tsai WB, Chen HK, Yam RSW, Herricks EE (2013) A self-organizing radial basis network forestimating riverine fish diversity. J Hydrol 476(1):280–289

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

Chitsazan N, Nadiri AA, Tsai FTC (2016) Prediction and structural uncertainty analyses of artificial neural networks using hierarchical Bayesian model averaging. J Hydrol 528:52–62

Cho KH, Sthiannopkao S, Pachepsky YA, Kim KW, Kim JH (2011) Prediction of contamination potential of groundwater arsenic in Cambodia, Laos, and Thailand using artificial neural network. Water Res 45(17):5535–5544

Chowdhury M, Alouani A, Hossain F (2010) Comparison of ordinary kriging and artificial neural network for spatial mapping of arsenic contamination of groundwater. Stoch Environ Res Risk Assess 24(1):1–7

Cigizoglu HK (2003) Estimation, forecasting and extrapolation of flow data by artificial neural networks. Hydrol Sci J 48(3):349–361

Coppola EA Jr, Rana AJ, Poulton MM, Szidarovszky F, Uhl VW (2005) A neural network model for predicting aquifer water level elevations. Ground Water 43(2):231–241

Dar IA, Sankar K, Dar MA, Majumder M (2012) Fluoride contamination - Artificial neural network modeling and inverse distance weighting approach. J Water Sci 25(2):165–182

Ding S, Zhang J, Xu X, Zhang Y (2016) A wavelet extreme learning machine. Neural Comput Appl 27(4):1033–1040

Domenico PA, Schwartz FW (1990) Physical and Chemical Hydrogeology. Wiley, New York, p 824

Dongwen CUI (2013) Application of extreme learning machine to total phosphorus and total nitrogen forecast in lakes and reservoirs. Water Resour Prot 29(2):61–66. doi:10.3969/j.issn.1004-6933.2013.02.013

Ebtehaj I, Bonakdari H, Shamshirband S, Mohammadi K (2016) A combined support vector machine-wavelet transform model for prediction of sediment transport in sewer. Flow Meas Instrum. doi:10.1016/j.flowmeasinst.2015.11.002

EPA (1997) Public health global for fluoride in drinking water. Pesticide and environmental toxicology. Section Office of Environmental Health Hazard Assessment, California Environmental Protection Agency

Fernández-Delgado M, Cernadas E, Barro S, Ribeiro J, Neves J (2014) Direct kernel perceptron (DKP): ultra-fast kernel ELM-based classification with noniterative closed-form weight calculation. Neural Netw 50:60–71

Fernando DAK, Shamseldin AY (2009) Investigation of internal functioning of the radial-basis-function neural network river flow forecasting models. J Hydrol Eng 14(3):286–292

Fijani E (2007) Investigation of hydrogeology and hydrochemistry of basaltic-alluvial aquifers in Bazarganand Poldasht Plains, MSc Thesis. University of Tabriz, Iran. (in Persian).

Fijani E, Nadiri A, Asghari Moghaddam A, Tsai FT-C, Dixon B (2013) Optimization of DRASTIC method by supervised committee machine artificial intelligence to assess groundwater vulnerability for Maragheh-Bonab plain aquifer. Iran. J Hydrol 503:89–100

Goyal MK, Bharti B, Quilty J, Adamowski J, Pandey A (2014) Modeling of daily pan evaporation in sub-tropical climates using ANN, LS-SVR, Fuzzy Logic, and ANFIS. Expert Syst Appl 41:5267–5276

Gupta S, Banerjee S, Saha R, Datta JK, Mondal N (2006) Fluoride geochemistry of groundwater in Birbhum, West Bengal, India. Fluoride 39:318–320

Haykin S (1999) Neural networks: a comprehensive foundation. Prentice-Hall, New Jersey, p 842

Huang G, Huang GB, Song S, You K (2015) Trends in extreme learning machines: a review. Neural Netw 61:32–48

Huang G, Song S, Gupta J, Wu C (2014) Semi-supervised and unsupervised extreme learning machines. IEEE Trans Cybern 44:2405–2417

Huang G-B, Zhou H, Ding X, Zhang R (2012) Extreme learning machine for regression and multiclass classification. IEEE Trans Syst Man Cybern 42(2):513–529

Huang GB, Zhu QY, Siew CK (2004) Extreme learning Machine: a new learning scheme of feedforward neural networks. Int Joint Conf Neural Netw 2:985–990

Huang G-B, Zhu Q-Y, Siew C-K (2006) Extreme learning machine: theory and applications. Neurocomputing 70(1):489–501

Imen S (2015) Drinking water infrastructure assessment with teleconnection signals, satellite data fusion and mining. University of Central Florida, Orlando, p 149

Jain A, Srinivasulu S (2004) Development of effective and efficient rainfall runoff models using integration of deterministic, real-coded genetic algorithms, and artificial neural network techniques. Water Resour Res 40(4):W04302. doi:10.1029/2003WR002355

Javadi A, Al-Najjar M (2007) Finite element modeling of contaminant transport in soils including the effect of chemical reactions. J Hazard Mater 143(3):690–701

Kisi O, Tombul M, Zounemat Kermani M (2015) Modeling soil temperatures at different depths by using three different neural computing techniques. Theor Appl Climatol 121(1):377–387

Kraynov SR, Merkov AN, Petrova NG, Baturinskaya IV, Zharikova VM (1969) Highly alkaline (pH 12) fluosilicate waters in the deeper zone of the Lovozero Massif. Geochem Int 6:635–640

Kundu MC, Mandal B (2009) Assessment of potential hazards of fluoride contamination in drinking groundwater of an intensively cultivated district in West Bengal, India. Environ Monit Assess 152:97–103

Liu Q, Yin J, Leung VCM, Zhai JH, Cai Z, Lin J (2016) Applying a new localized generalization error model to design neural networks trained with extreme learning machine. Neural Comput Appl 27:59–66

Malekmohammadi I, Bazargan-Lari MR, Kerachian R, Nikoo MR, Fallahnia M (2011) Evaluating the efficacy of SVMs, BNs, ANNs and ANFIS in wave height prediction. Ocean Eng 38:487–497

Mirabbasi R (2015) Application of artificial intelligence methods for groundwater quality prediction. Application of artificial intelligence methods in geosciences and hydrology. MICS Group International, hyderabad

Mishra AK, Desai VR (2006) Drought forecasting using feed-forward recursive neural network. Ecol Model 198(1–2):127–138

Mojumder JC, Ong HC, Chong WT, Shamshirband S, Al-Mamoon A (2016) Application of support vector machine for prediction of electrical and thermal performance in PV/T system. Energy Build 111:267–277

Nadiri AA, Fijani E, Tsai FTC, Asghari Moghaddam A (2013) Supervised committee machine with artificial intelligence for prediction of fluoride concentration. J Hydroinform 15(4):1474–1490

Nash JE, Sutcliffe JV (1970) River flow forecasting through conceptual models part I—a discussion of principles. J Hydrol 10(3):282–290

Noori R, Karbassi AR, Moghaddamnia A, Han D, Zokaei-Ashtian MH, Farokhnia A, Ghafari Gousheh M (2011) Assessment of input variables determination on the SVM model performance using PCA, Gamma test, and forward selection techniques for monthly stream flow prediction. J Hydrol 401:177–189

Noori R, Karbassi AR, Sabahi MS (2010) Evaluation of PCA and Gamma test techniques on ANN operation for weekly solid waste predicting. J Environ Manag 91:767–771

Nourani V, Baghanam AH, Adamowski J, Gebremichael M (2013) Using self-organizing maps and wavelet transforms for space–time pre-processing of satellite precipitation and runoff data in neural network based rainfall-runoff modeling. J Hydrol 476:228–243

Patel SC, Khalkho R, Patel SK, Sheikh JM, Behera D, Chuadhari S, Prabhakar N (2014) Fluoride contamination of groundwater in parts of eastern India and a preliminary experimental study of fluoride adsorption by natural haematite iron ore and synthetic magnetite. Environ Earth Sci 72:2033–2049

Rafique T, Naseem S, Bhanger MI, Usmani TH (2008) Fluoride ion contamination in the groundwater of Mithi sub-district, the Thar Desert, Pakistan. Environ Geol 56:317–326

Rajasekaran S, Gayathri S, Lee TL (2008) Support vector regression methodology for storm surge predictions. Ocean Eng 35(16):1578–1587

Ramanaiah SV, Venkatamohan S, Rajkumar B, Sarm PN (2006) Monitoring of fluoride concentration in groundwater of Prakasham district in India: correlation with physico-chemical parameters. J Environ Sci Eng 48:129–134

Sahoo GB, Ray C, Mehnert E, Keefer DA (2006) Application of artificial neural networks to assess pesticide contamination in shallow groundwater. Sci Total Environ 367(1):234–251

Shamshirband S, Mohammadi K, Khorasanizadeh H, Yee PL, Lee M, Petkovic D, Zalnezhad E (2016) Estimating the diffuse solar radiation using a coupled support vector machine–wavelet transform model. Renew Sustain Energy Rev 56:428–435

Sirat M (2013) Neural network assessment of groundwater contamination of US Mid-continent. Arabian Journal of Geosciences 6(8):3149–3160

Sonebi M, Cevik A, Grunewald S, Walraven J (2016) Modelling the fresh properties of self-compacting concrete using support vector machine approach. Constr Build Mater 106:55–64

Valenzuela-Vasquez L, Ramýrez-Hernandez J, Reyes-Lopez J, Sol-Uribe A, Lazaro-Mancilla O (2006) The origin of fluoride in groundwater supply to Hermosillo City, Sonora, Mexico. Environ Geol 51:17–27

Vapnik VN (1995) The nature of statistical learning theory. Springer, New York, p 314

Wanas N, Auda G, Kamel MS, Karray F (1998) On the optimal number of hidden nodes in a neural network. Proc IEEE Can Conf Electr Comput Eng 2:918–921

Wang B, Tang L, Yang J, Zhao B, Wang S (2015) Visual tracking based on extreme learning machine and sparse representation. Sensors 15(10):26877–26905

World Health Organization (2008) Guidelines for drinking water quality. Recommendations, 2nd edn. WHO, Geneva, p 306

Wu KP, Wang SD (2009) Choosing the kernel parameters for support vector machines by the inter-cluster distance in the feature space. Pattern Recogn 42(5):710–717

Yang H, Huang K, King I, Lyu MR (2009) Localized support vector regression for time series prediction. Neurocomputing 72(10):2659–2669

Zaji AH, Bonakdari H, Khodashenas SR, Shamshirband S (2016) Firefly optimization algorithm effect on support vector regression prediction improvement of a modified labyrinth side weir's discharge coefficient. Appl Math Comput 274:14–19

Zhang Y, Cai Z, Gong W, Wang X (2015) Self-adaptive differential evolution extreme learning machine and its application in water quality evaluation. J Comput Inf Syst 11(4):1443–1451