Heliyon 9 (2023) e18415

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Viability of two adaptive fuzzy systems based on fuzzy c means and subtractive clustering methods for modeling Cadmium in groundwater resources

Naghmeh Jafarzade [a], Ozgur Kisi [b,c], Mahmood Yousefi [d], Mansour Baziar [e], Vahide Oskoei [f], Nilufar Marufi [g], Ali Akbar Mohammadi [h,*]

[a] Department of Natural Resources and Environment, Science and Research Branch, Islamic Azad University, Tehran, Iran
[b] Department of Civil Engineering, Technical University of Lübeck, 23562, Lübeck, Germany
[c] Department of Civil Engineering, Ilia State University, 0162, Tbilisi, Georgia
[d] Department of Environmental Health Engineering, School of Public Health, Iran University of Medical Sciences, Tehran, Iran
[e] Department of Environmental Health Engineering, Ferdows Faculty of Medical Sciences, Birjand University of Medical Sciences, Birjand, Iran
[f] School of Life and Environmental Science, Deakin University, Geelong, Australia
[g] Department of Environmental Health Engineering, School of Public Health and Safety, Shahid Beheshti University of Medical Sciences, Tehran, Iran
[h] Department of Environmental Health Engineering, Neyshabur University of Medical Sciences, Neyshabur, Iran

A R T I C L E I N F O

Keywords:
Heavy metal prediction
Cadmium
Groundwater
Adaptive fuzzy systems
Fuzzy clustering

A B S T R A C T

The Adaptive Neuro-Fuzzy Inference System (ANFIS) combines the strengths of both Artificial Neural Networks (ANNs) and Fuzzy Logic (FL) into a single framework. By doing so, it allows for quicker learning and adaptable interpretation capabilities, which are useful for modeling complex patterns and identifying nonlinear relationships. One significant challenge in assessing water quality is the difficulty and time-consuming nature of determining the various factors that impact it. Given this situation, predicting heavy metal levels in groundwater resources, both urban and rural, is essential. This paper investigates two methods, ANFIS-FCM and ANFIS-SUB, to determine their effectiveness in modeling Cadmium (Cd) in groundwater resources. The parameters to be considered are: dissolved solids (TDS), electroconductivity (EC), turbidity (TU), and pH were assumed to be the independent variables. A total of 51 sampling location were used with in the groundwater resource were used to develop the fuzzy models. For evaluating the performance of ANFIS-FCM and ANFIS-SUB models, three different performance criteria including the correlation coefficient, root mean square error, and sum square error were used for comparing the model outputs with actual outputs. Based on the obtained results from scatter plots of actual and predicted value by ANFIS-SUB and ANFIS- FCM models, the determination coefficient (R^2) value for total data, test and train sets is equal to 0.978, 0.982, 0.993 and to 0.983, 0.999 and 0.998 respectively. This result proved the Cd predictions of the implemented ANFIS-FCM model was significantly close to the measured all experimental data with R^2 of 0.983. The performance of the implemented ANFIS-FCM model was compared with the ANFIS-SUB model and it is found that the ANFIS-FCM provided slightly higher accuracy than the ANFIS-SUB model. Also, the results obtained from the comparison between the predicted and the actual data indicated that the ANFIS-FCM and ANFIS-SUB have a strong potential in estimating the heavy metals in the groundwater with a high degree of accuracy.

* Corresponding author.
E-mail address: Mohammadi.eng73@gmail.com (A.A. Mohammadi).

https://doi.org/10.1016/j.heliyon.2023.e18415
Received 12 January 2023; Received in revised form 16 July 2023; Accepted 17 July 2023
Available online 18 July 2023

1. Introduction

In developing countries, the rapid growth of industries such as petrochemicals, and heavy automotive industries, along with strong economic growth and the unprecedented expansion of cities, causes a significant increase in the emission of pollutants that include environmental pollution, which is considered as a national and international issue [1,2]. As a result of these activities, pollutants such as toxic metals, hydrocarbons, priority organic pollutants, pesticides, nanoparticles, microplastics, and other emerging pollutants have the possibility of entering the groundwater resources and are considered a threat to human health, environmental services and sustainable social and economic development [3–5]. Among the mentioned pollutants, metals and toxic metalloids are risk factors for the health of the human population and the natural environment [6–10]. Heavy metals (HMs.) pollution is a global challenge that has very important effects on the health of society and ecological species [11]. In recent decades, studies conducted in different parts of the world have reported contamination of groundwater resources with As, lead, iron, manganese, cadmium, copper, and chromium [12–16]. Contamination of soil and groundwater with cadmium is a global problem affecting food and drinking water supplies mainly in Asia and Africa [17]. Cadmium (Cd) is a highly toxic HM, even at low concentrations. It leaches into the soil by water and further bio-accumulates in ecosystems and organisms. Long-term ingestion of Cd contaminated water can cause a variety of diseases, such as liver, kidneys, immune system, bones, and reproductive organs, anemia, cancer, and cardiovascular problems [17,18]. Therefore, knowing the quality of drinking water sources is always a necessary and inevitable thing to maintain the health of society, which requires different stages of water sampling, analyzing, providing reagents, chemicals, and calibration standard solutions and lack of skilled staffs. According to the above, the researchers looked for new approaches that can easily perform the assessment and monitoring, so that these problems in sampling and finally monitoring, evaluating, and management of the water resources can be minimized [19]. Therefore, it seems necessary to use innovative tools and methods to address the above issues regarding environmental quality monitoring, especially in areas where there is a possibility of contamination of water resources [20]. Machine learning (ML) is an effective tool for extracting predictive models from data. Today, ML are widely used in many fields, including environmental modeling, water resources engineering, and predicting climate change, environmental pollution and forecasting concentration of heavy metals [21–24]. Modeling complex nonlinear systems is one of the successful applications of artificial intelligence (AI)-based techniques, such as fuzzy inference systems (FIS), artificial neural network (ANN), and genetic algorithms [25,26]. ANN is widely used in various fields of environmental engineering such as monitoring of groundwater and surface water resources and air quality monitoring. Although the modeling of heavy metals, including cadmium, has been studied using various statistical and computational methods [10,27,28]. In the early 2020s, ANFIS techniques became popular among researchers and have since been utilized for numerous forecasting applications [29–34]. For instance, Dus (2018) used four techniques including Backpropagation neural network (BPNN), Radial basis function network (RBFN), Recurrent neural network (RNN), and Adaptive Neuro-Fuzzy Inference System (ANFIS) to predict the aquifer potential of a well [35]. Samantaray (2022) explored the potential of three machine learning models, namely radial basis function network (RBFN), support vector machine (SVM), and integration of SVM with firefly algorithm (SVM-FFA), to forecast GWL of two subwatersheds in Nuapada district, Odisha, India. Their study showed that the results of ANFIS-GWO models outperformed standalone ANFIS and conventional CFBPNN models [31,34]. Emamgholizadeh (2023) predicted soil cation exchange capacity using enhanced machine learning approaches in the southern region of the Caspian Sea. The study found that coupled models (ANFIS-DE and ANFIS-PSO) were more efficient than the ordinary ANFIS model [36]. Hadadi (2022) applied classic ANFIS to estimate the daily dew point temperature and found that hybrid models (i.e., ANFIS-BCO and ANFIS-DFA) demonstrated better performance compared to classic ANFIS [37]. Several researchers have conducted studies on various ANFIS and AI techniques for water quality monitoring and assessment, including the prediction of heavy metal concentrations. Ahmed et al. (2015) estimated the biochemical oxygen demand (BOD) of Surma River [38]. AkpoFmie et al. (2016) modeled the concentration of heavy metals in artificial borings

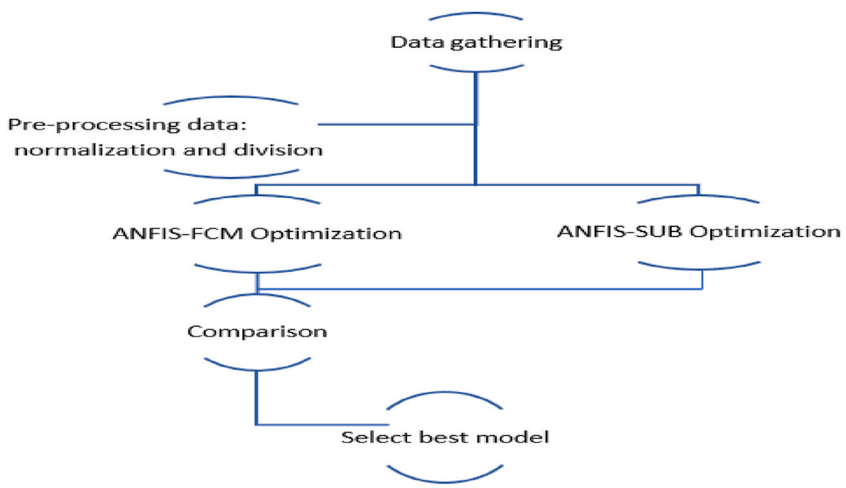

Fig. 1. Flow chart of the developed model.

[39], while Areerachakul et al. compared ANFIS and ANN for the estimation of biochemical oxygen demand in surface water [40]. Chang et al. used neuro-fuzzy networks to estimate the prediction of arsenic concentration in stream water [41]. Dahiya et al. applied fuzzy synthetic to assess groundwater quality [42], and Yılmaz et al. analyzed the application of fuzzy logic for the estimation of heavy metal pollution in Apa Dam Lake [43]. Given that around 90% of the required water for Neyshabur Plain's domestic and agricultural purposes is supplied by groundwater resources, this study was conducted to predict the cadmium content in the groundwater resources of Neyshabur Plain. For this purpose we evaluate the viability of two adaptive fuzzy systems based on fuzzy C-means and subtractive clustering methods in modeling Cadmium in the Neyshabur groundwater resource according developed model (Fig. 1).

2. Material and methods

2.1. Study area and sampling

Neyshabur city is one of the central parts of Razavi Khorasan province, which now has an area of 5805 square kilometers, between latitudes 35° and 34 min to 36° and 56 min north and longitude 58° and 10 min to 58° and 62 min east. It is located in the central desert of Iran (Fig. 2). Most of the city is located on a relatively wide plain, which is limited from the north and east by the heights of Binalud. Therefore, the most important feature of the geographical location of Neyshabur is the extension of high mountain ranges around it and the existence of mountain views in most parts of this city. The climate of the region is semi-arid to dry and the average monthly temperature of Bar station (representing mountainous areas) is 13° Celsius and Mohammad Abad-Fadisheh station (representing plain areas) is 13.8° Celsius. Despite the small temperature difference between the highlands and the plain, the climate of the basin in the north and the south is very different, so that in the north, which is mountainous, the weather is relatively cold with a mild summer, and to the south and west, the weather becomes warmer, which can be caused by the large area. Be a basin The average rainfall in the whole basin is equal to 234 mm, although the amount of rainfall is different at different points so that in the high altitudes of Binaloud, its amount is up to 600 mm and in the plains, it is many times less than that in the winter season [44].

2.2. Data gathering

A total of 158 water samples were taken from all the wells in the area that were used for drinking water during 2018. Sampling was done according to the standard method using 1.5-L polyethylene containers, then it was taken to the laboratory at a temperature of 4° Celsius. Sampling was done instantaneously in sterile polyethylene containers free from pollution and after 10–20 min of pumping water from the well outlet. Sampling and measurement of quality parameters were done for water and wastewater based on standard methods [45]. The qualitative parameters used in this study include electrical conductivity (EC), acidity (pH), total dissolved solids (TDS), and turbidity. The above parameters were measured by EC meter, pH meter, and colorimeter also Cd concentrations were analyzed by inductively coupled plasma optical emission spectrometry. The box plot of raw data used in this study is shown in Fig. 3. This chart is a standard way to display data distribution and can provide information about outliers. It also provides information about data compression or symmetry.

2.3. Adaptive Neuro-Fuzzy Inference System (ANFIS)

ANFIS was first introduced by J.S. Range in 1993 [46]. ANFIS is a natural network that works similarly to the model of

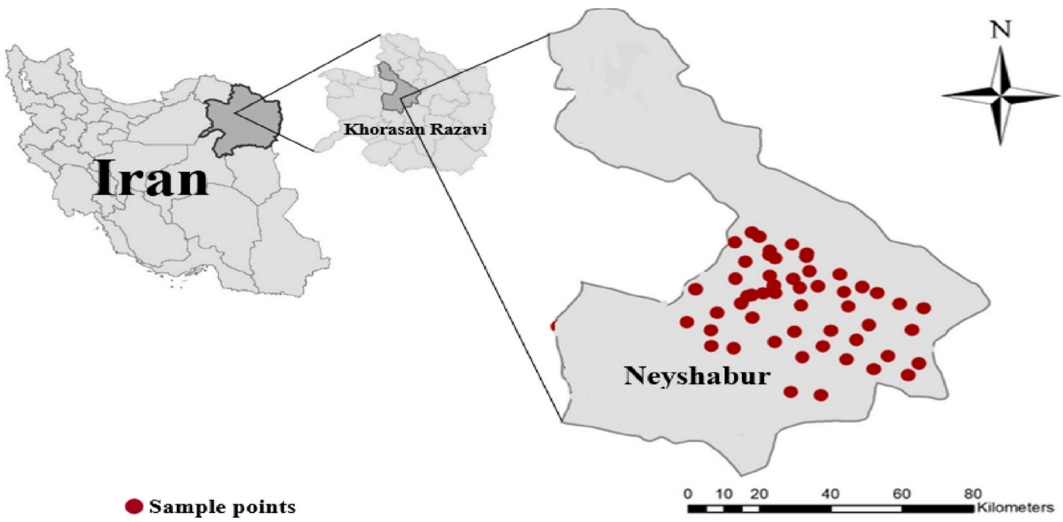

Fig. 2. Map of the study area in Iran.

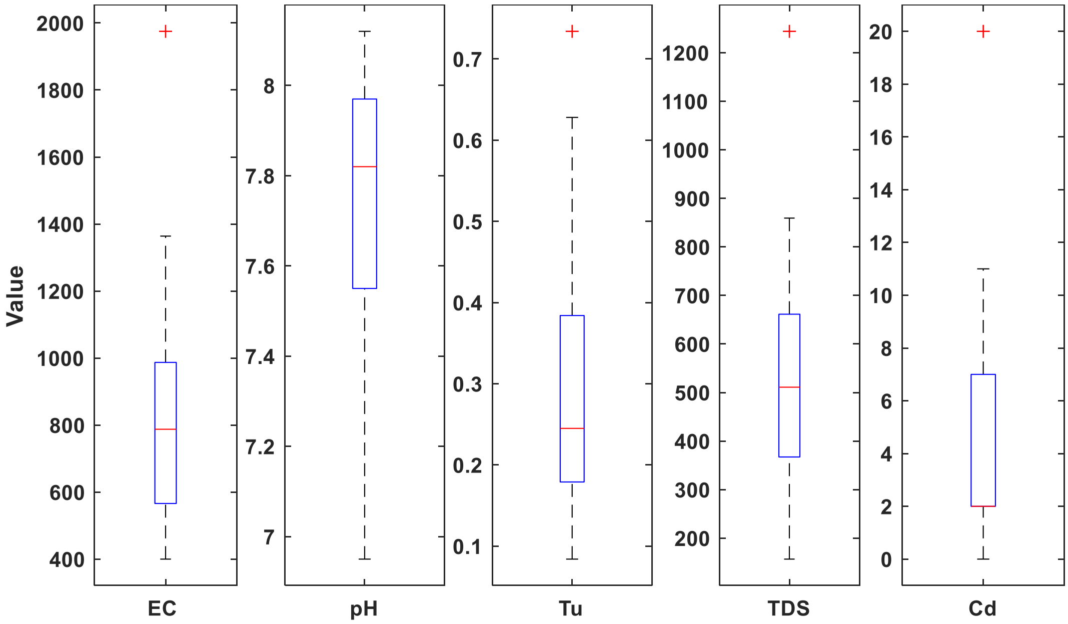

Fig. 3. Box plot of raw data used in this work.

Takagi–Sugeno model. The ANFIS, as a hybrid intelligent system, uses ANN and fuzzy logic together in a unified system. So, critical problems of designing fuzzy rules are easy to solve by ANFIS, and it develops the rules and parameter optimization by using the learning capability of the ANN. The mechanism of working in this unified system is as the ANFIS could learn from the prepared training data by the ANN techniques, which has a duty to update the Takagi–Sugeno inference model parameters. Hence, the Fuzzy Interface System (FIS) as an intercessor device produces the designed results in the type of linguistic terms. The structure of the ANFIS model is mainly described by utilizing five separate specific layers to explain the model concept (Fig. 4). According to the hierarchy and their role, these five layers comprise the fuzzification layer (layer 1); any node of this layer is expected to be the adaptive node. Layer 1 outputs are a degree of membership allocated to inputs in the fuzzy forms, the layer acting as the rule base layer (layer 2), which is

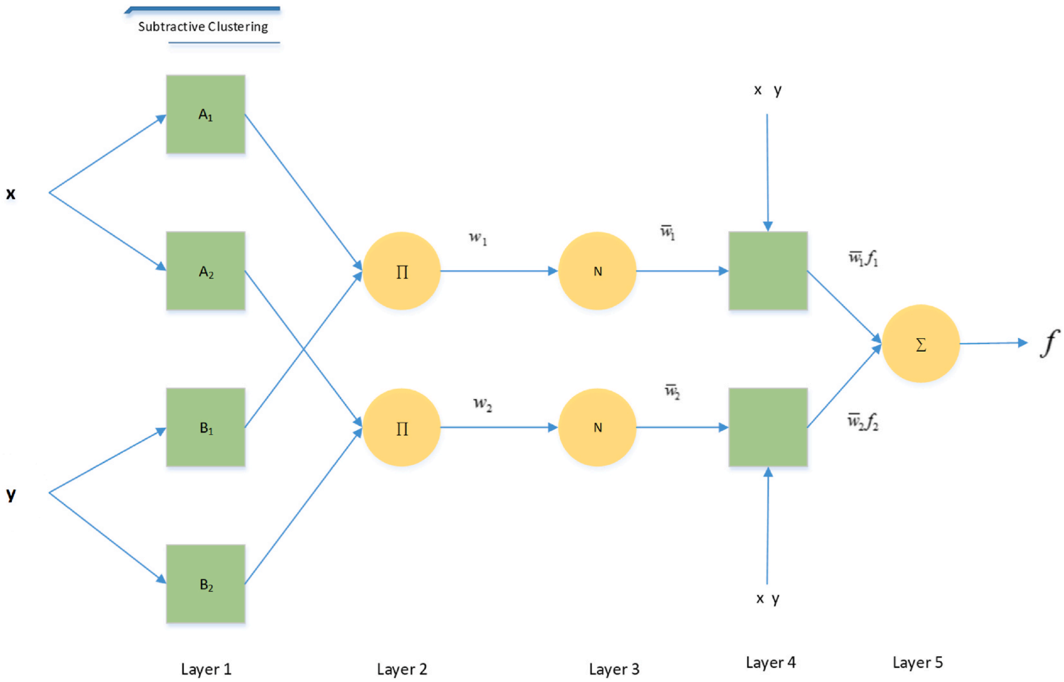

Fig. 4. ANFIS structure with two inputs and one output and with subtractive clustering.

famous for the firing strength of the rules. The nodes of this layer are the fixed ones and nominated with Π, layer 3 (normalization of the membership functions), nodes of this layer are set like the former layer node and nominated with N. Any node depicts each rule's normalized firing strength. Layer 4, nominated as defuzzification one, contains adaptive nodes with node behavior. Layer 5, called the summation layer (a single node) nominated by Σ, illustrates the final outputs as the incoming signals' summation.

The designed structure for the ANFIS model has been defined as a network containing two 'Constructing' and 'Training' sections. The ANFIS's construction section is a place that determines the MFs' types and numbers. The first necessity of this section is that the input and output data are subdivided into rule patches. Diverse types of methods achieve these necessities, comprising fuzzy c-means (FCM), subtractive clustering, and grid partitioning [47].

2.4. Subtractive clustering

The rules of IF-THEN determine the TS fuzzy modeling method, but it has some limits because of the manual inspection due to the incapability of recognizing all the rules. Therefore, it is better to use fuzzy clustering techniques that help identify the rules by utilizing the recorded data in that situation. Fuzzy clustering is primarily utilized to identify and classify indistinguishable patterns from massive datasets into numerous categories. That could find the number of expected clusters by users, or it is possible to vest the system to find the feasible number of clusters that can be extractable from the input data [26]. Various fuzzy clustering methods have been suggested in the former research in which fuzzy C-means clustering, subtractive clustering, and mountain clustering are nominated as the famous ones. The subtractive clustering method is mainly utilized to prepare clues for various sciences and difficulties, referred to the engineering [22,47]. In this mentioned method, the foremost belief is that all original data points (Fig. 3) possess the capability to perform as cluster centers. So, each data point's rank is measured according to the density of the surrounding ones. This score could be the addition of the distance data point to the specific point. A cluster center was chosen where any data point took a fewer score (Fig. 3/orange-colored node). The later cluster and its information center are determined by discarding the influenced data points with radios (Fig. 3/evident from purple-colored node). This procedure has a continuous manner till a specific number of clusters is developed. Fig. 5 illustrates the different stages of subtractive clustering implementation.

2.5. Fuzzy C-means (FCM)

FCM is used for data clustering where an identified dataset is divided into some clusters based on the fuzzy C-partition's principles. Firstly, FCM was introduced by Ruspini and then developed and generated by Dunn and Bezdek [48,49]. In this method, any dataset's point is a part of a cluster constant of the score of their membership degree. For example, data points close to the cluster center have the

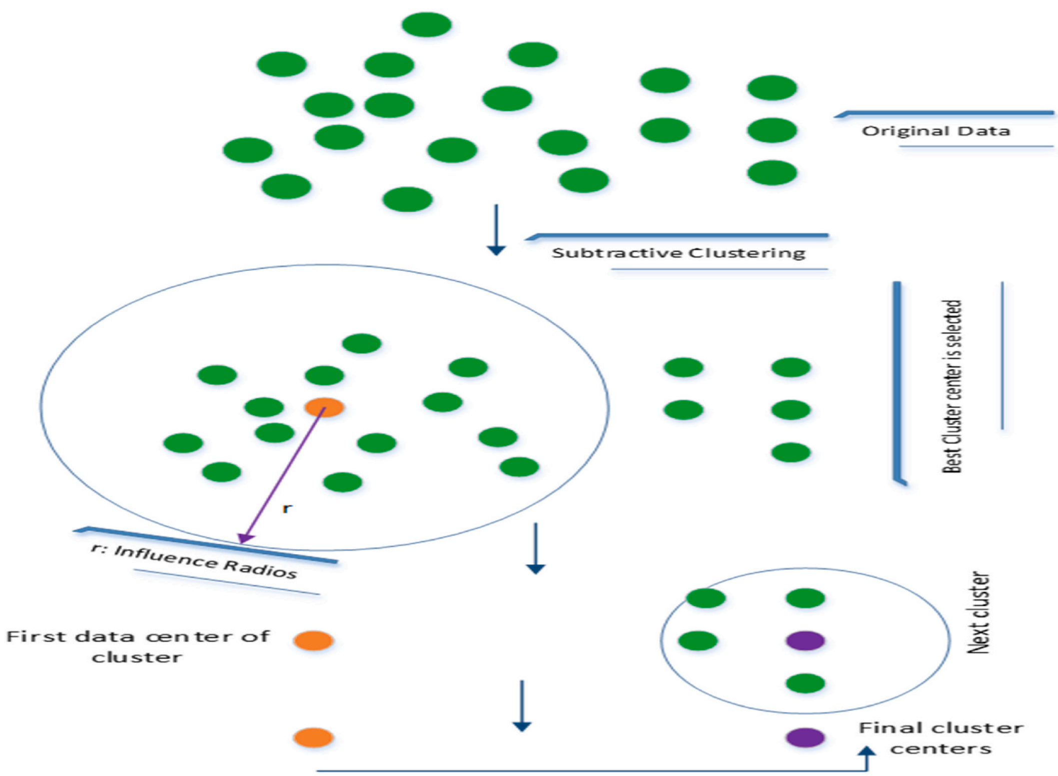

Fig. 5. Subtractive clustering method.

highest belonging score (membership degree) to that cluster, and others so far from the center have the lowest belonging score. Generally, inserting every data point into one of the clusters is the critical aim of the FCM. This was conducted using the FCM method to separate the training dataset into different numbers of subsets (Gaussian membership functions) with various centers. Each subset is trained by utilizing ANFIS to detect the foremost number of membership functions (MFs) according to the least MSE in the training, validation, and testing stages [50].

2.6. Data processing

The experimental datasets in this research were casually subdivided into training (80%) and test (20%) subsets before the development of the ANFIS-FCM models. The experimental data was divided with the aim of evaluating the implemented models' robustness in predicting unseen data. Furthermore, inputs (process parameters) and outputs (Cd concentrations) of the ANFIS-FCM and ANFIS – SUB models were normalized in the range of 0.1 and 0.9, using the relation of Eq. (1).

$$y = \frac{x_i - x_{min}}{x_{max} - x_{min}} \times (b - a) + a \tag{1}$$

here, y nominated as the normalized value of x_i, x_{min} and x_{max} are maximum and minimum values of the experimental data [49,50].

3. Results

In this study for evaluating the performance of ANFIS-FCM and ANFIS-SUB models, three different performance criteria including the correlation coefficient (R^2), root mean square error (RMSE), and sum square error (SSE) were used for comparing the model predicted outputs with actual outputs. The Cd concentration and input variables including in-situ physicochemical parameters pH, TDS, EC, and turbidity were measured at 54 stations wells, springs, and subterranean in Neyshabur city.

3.1. ANFIS-FCM results

The utilization of fuzzy inference systems with various functions has been expanded due to the capability of solving engineering problems, its ability in control systems, and a clear definition of variables in terms of linguistic ones. Although a robust fuzzy inference system design is not easy to access and requires determining appropriate membership functions and fuzzy rules, it is accessible with so much trial-error and the employment of experts' comments to obtain superior accuracy. However, sometimes, it is hard or impossible to identify the rules. Therefore, ANN learning algorithms are used for fuzzy systems. The ANN part of the ANFIS has two learning strategies: Hybrid learning and back-propagation. In the fuzzy section of the ANFIS, exclusively the zero or first-order Sugeno fuzzy inference system might be utilized [31,32].

The construction of ANFIS-FCM to predict Cd concentration is illustrated in Fig. 5. According to the figure, the output (Cd concentration) is measured by employing 10 fuzzy rules to fuzzy sets of four inputs: pH, TDS, EC, and turbidity. The working mechanism of an ANFIS is the application of the noticed ANN learning attitude to accommodate the FIS parameters. The original FIS that includes the Gaussian MF was defined by FCM for any variables. The MATLAB code was created with the assistance of MATLAB R-2013b software. The ANFIS-FCM models' performance was evaluated according to the MSE. Table 1 illustrates the outcomes of the various number of Gaussian MFs (cluster number), optimal method, epoch numbers, and values of MSE.

It was identified that if the ANFIS-FCM model with 10 memberships for every variable, the hybrid method and 200 epochs provided the MSE of 1.1E-09 and 0.0018 for training and testing datasets. It is evident from the best ANFIS-FCM (Fig. 6) that there are 10 rules to describe the Cd concentration.

The scatter plots of the observed and predicted Cd concentrations by the ANFIS-FCM in the total data, training and testing sets are illustrated in Fig. 7(A-C). The determination coefficient (R^2) value for total data, test and trains sets is equal to 0.983, 0.998 and 0.999

Table 1
The outcomes of the ANFIS-FCM in Cd prediction (normalized scale).

Input MF type	Number of cluster	Output MF type	Optimum method	Epoch	MSE	
					Training	Testing
gaussmf	2	Linear	Hybrid	100	2. 1E-03	0.1430
gaussmf	3	Linear	Hybrid	100	2*10-52.0E-05	0.0067
gaussmf	4	Linear	Hybrid	100	1.2E-09	0.0466
gaussmf	5	Linear	Hybrid	100	8.0E-10	0.0497
gaussmf	6	Linear	Hybrid	100	1.6E-10	0.0546
gaussmf	7	Linear	Hybrid	100	1.5E-09	0.0231
gaussmf	8	Linear	Hybrid	100	1.0E-10	0.0169
gaussmf	9	Linear	Hybrid	100	8.0E-10	0.0307
gaussmf	10	Linear	Hybrid	100	8.0E-10	0.0035
gaussmf	10	Linear	Back propagation	100	1.8E-03	0.0268
gaussmf	10	Linear	Back propagation	200	5.8E-04	0.0121
gaussmf	10	Linear	Hybrid	200	1.1E-09	0.0018

respectively. The training test had the highest R^2 value of 0.999, indicating a strong correlation between the actual and predicted values. The R^2 value illustrates that the designed model with 10 linguistic fuzzy rules could explain 98.3% variation between the predicted and target values. Fig. 8 demonstrates the residual errors' distribution which varies from −1.8 to 6.82E-06.

3.2. ANFIS-SUB results

The experimental dataset (input and output) was clustered in different numbers in the training period based on the subtractive clustering method. The proper model was achieved based on better prediction accuracy after training with the ANFIS model with 12 Gaussian membership functions for every single input. The ANFIS -subtractive clustering method generated 12 fuzzy rules for the best-designed model with a cluster radius of 0.4 (Table 2). So, the MSE values of the training and testing data sets were equal to 1.1E-10 and 0.002, accordingly. Fig. 9 demonstrates the predictions of ANFIS-SUB.

The scatter plots of the observed and predicted Cd concentrations by the ANFIS-SUB in the total data, testing and training sets are illustrated in Fig. 10(A-C). The determination coefficient (R^2) value for total data, test and train sets is equal to 0.978, 0.982 and 0.993 respectively. The highest R^2 value between the actual and predicted values is 0.993, which belongs to the training test.

In Figs. 9 and 10, the observed values and predicted outputs of the ANFIS-SUB and ANFIS-FCM models are compared. The training set for the ANFIS-FCM model has the highest R^2 value of 0.999 between actual and predicted values. However, when the testing set is applied, the R^2 value slightly decreases to 0.998. In general, the R^2 value of 0.983 for the total data is reported for the ANFIS-FCM model. These results indicate that the ANFIS-FCM algorithm is reliable for estimating Cd values. Overall, the generalization power of the ANFIS-SUB algorithm is slightly weaker than the ANFIS-FCM algorithm, and the predicted results are less consistent with the actual results. In the next stage of the study, a histogram association and mapping technique based on prediction errors were obtained from embedding the total data. The histogram errors for the total data indicate that the error value of ANFIS-FCM and ANFIS-SUB models ranges from −2 to 1 and -2 to 2, respectively as shown in Fig. 11A, B. ANFIS-FCM (Fig. 11 A) has shown the ability to provide predictions with smaller errors compared to ANFIS-SUB methods (Fig. 11B). Therefore, it can be suggested that ANFIS-FCM has demonstrated an acceptable performance in simulating data.

3.3. Comparison of two adaptive neuro fuzzy systems

Three assessment criteria comprising determination coefficient (R^2), root means square error (RMSE), and sum square error (SSE) were used in this research to measure ANFIS-FCM and ANFIS-SUB models' performance in predicting Cd concentration. The following equations refer to the assessment criteria (Eqs. (2)–(4)).

$$SSE = \sum_{i=1}^{n} (act_i - pre_i)^2 \quad (2)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{n} (act_i - pre_i)^2} \quad (3)$$

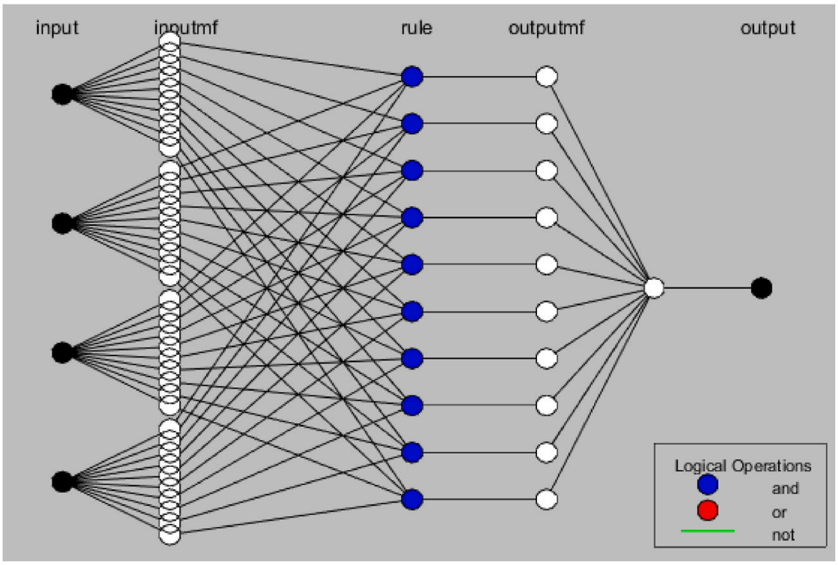

Fig. 6. The structure of the best-developed ANFIS-FCM model.

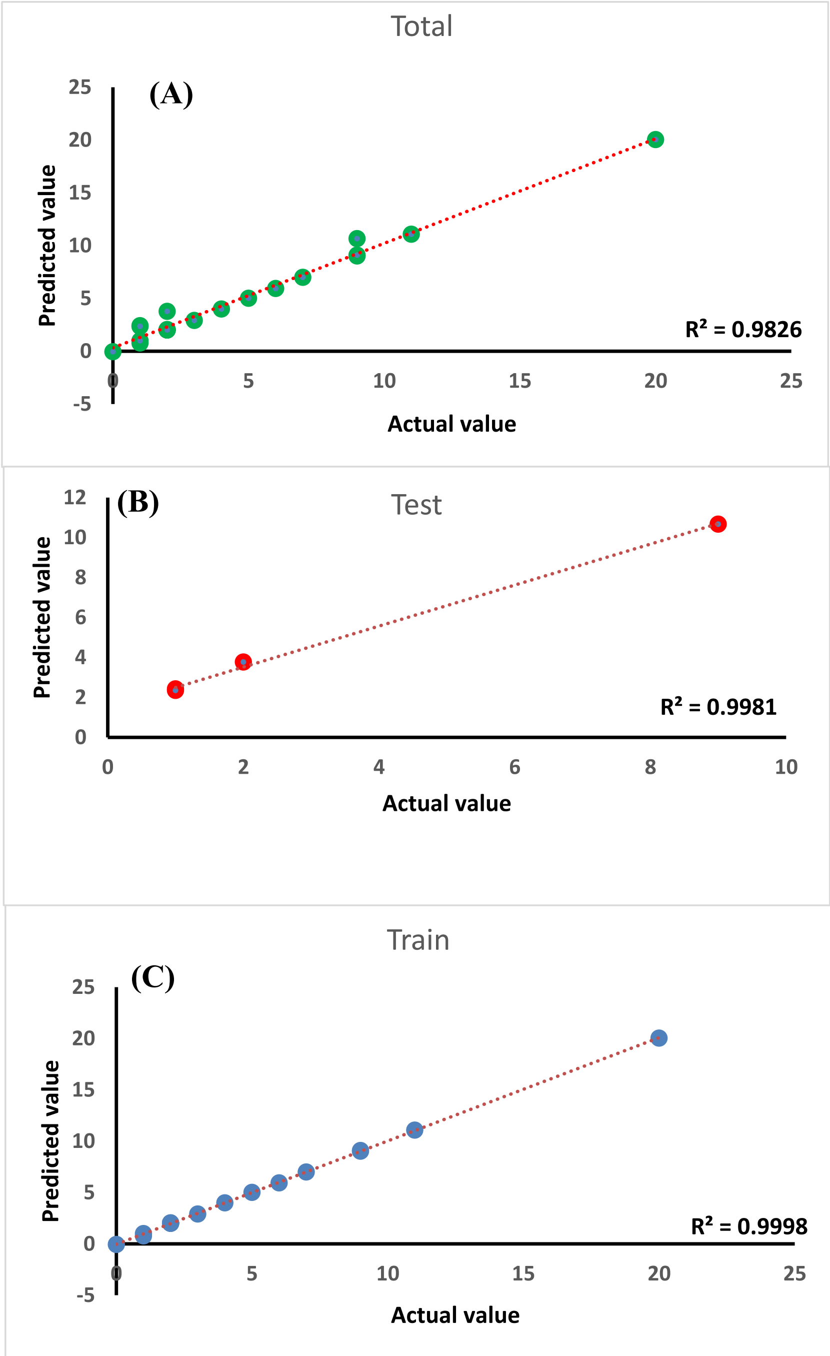

Fig. 7. Scatter plots of actual value and predicted value using (ANFIS-FCM): (A) total data(B) testing set and (C) Training set.

$$R^2 = \left(\frac{\sum_{i=1}^{n} (act_i - \overline{act})(pre_i - \overline{pre})}{\sqrt{\sum_{i=1}^{n} (act_i - \overline{act})^2 \sum_{i=1}^{n} (pre_i - \overline{pre})^2}} \right)^2 \tag{4}$$

Here, act_i is nominated as the actual value, pre_i is nominated as the predicted value, $\overline{act}$ and $\overline{pre}$ are the average values of actual values and predicted values, respectively. Table 3 illustrates the comparative performance of two neuro-fuzzy models and it is evident that the ANFIS-FCM performs slightly better than the ANFIS-SUB model [49,51].

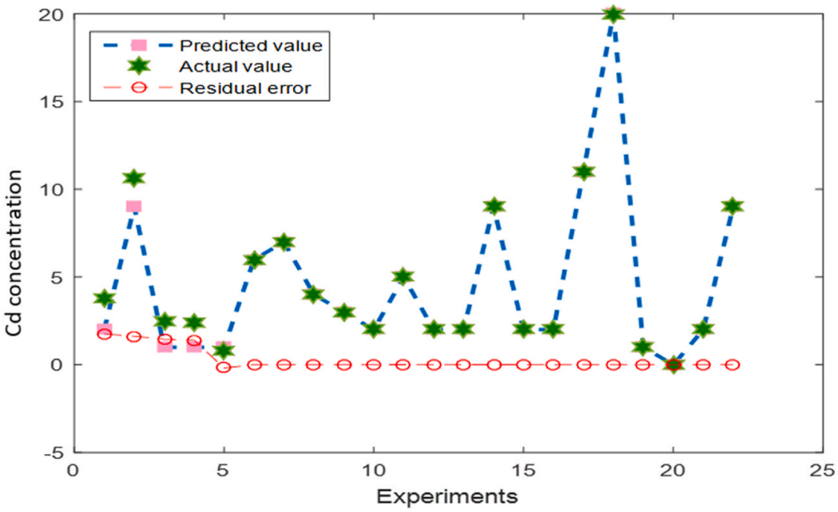

Fig. 8. The distribution of the residual errors for ANFIS-FCM model.

Table 2
ANFIS-SUB performance for different cluster radius (normalized scale).

Number of Cluster	Radius rules	Performance indices	
		MSE train	MSE test
0.1	17	3.6E-14	0.022
0.2	16	5.5E-11	0.006
0.3	16	7.6E-14	0.0524
0.4	12	1.1E-09	0.002
0.5	11	1.16E-10	0.0061
0.6	8	1.2E-10	0.0097
0.7	6	7.7E-09	0.07
0.8	5	3.2E-09	0.0522
0.9	4	5.8E-09	0.0508
0.95	4	1.6E-08	0.2335

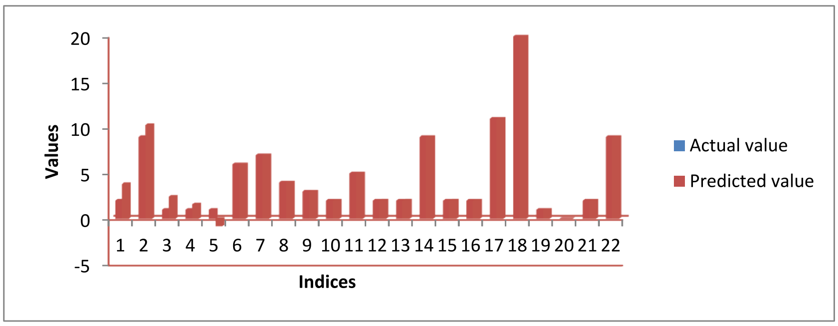

Fig. 9. The predictions of ANFIS-SUB method.

4. Discussion

Many studies have applied artificial intelligence-based models to predict the concentration of heavy metals in aquatic environments. The outcomes revealed that the Physicochemical properties of water resources have an effect on the concentration of heavy metals. In this study, three different performance criteria including the correlation coefficient (R^2), root mean square error (RMSE), and sum square error (SSE) were used for evaluating the performance of ANFIS-FCM and ANFIS-SUB models.

The estimates of ANFIS-FCM model showed that the determination coefficient (R^2) value for total data, test and train sets is equal to 0.983, 0.998 and 0.999 respectively. The highest R^2 value between the actual and predicted values is 0.999, which belongs to the training set. The R^2 value 0.983 for whole data illustrates that the designed model with 10 linguistic fuzzy rules could explain 98.3%

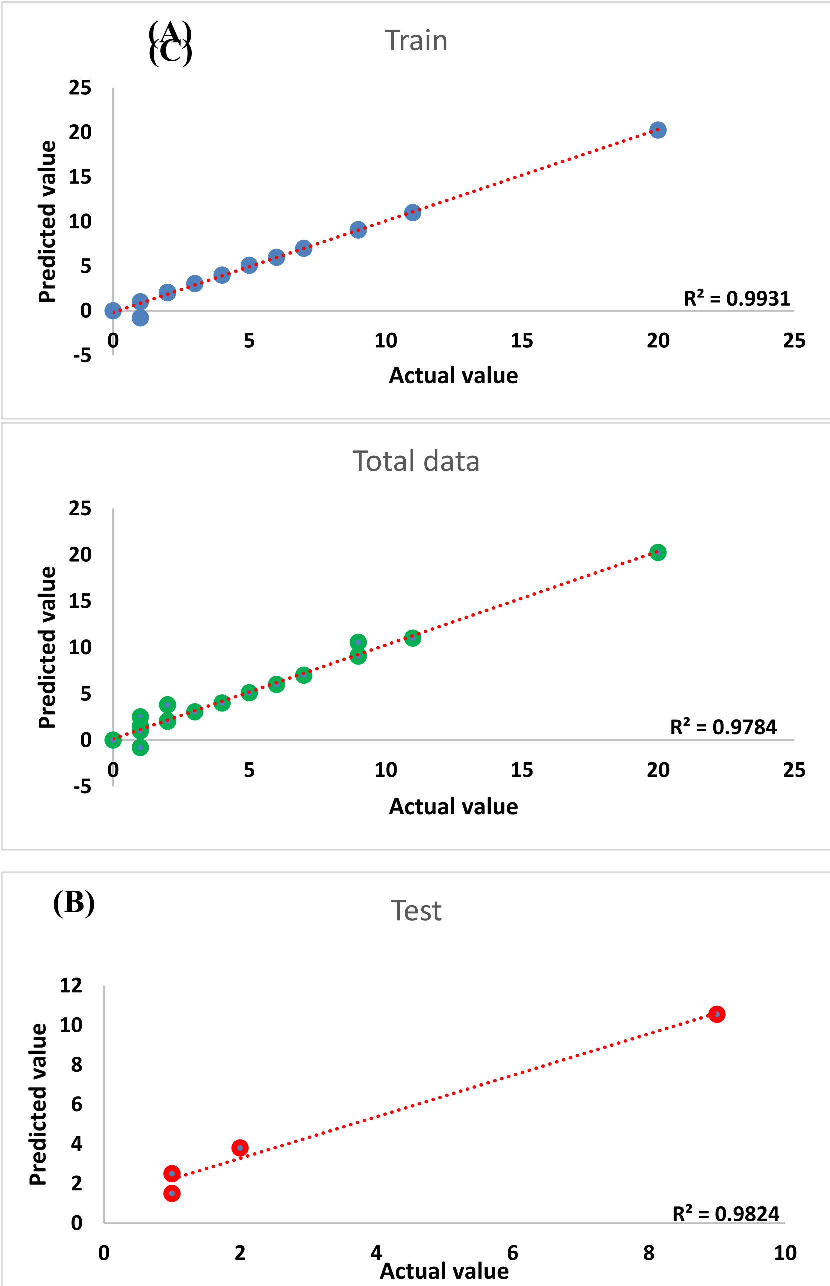

Fig. 10. Scatter plots of actual value and predicted value using (ANFIS-SUB): (A) total data(B) testing set and (C) Training set.

variation between the predicted and target values. In addition, the determination coefficient (R^2) value for total data, test and train sets according ANFIS-SUB model is equal to 0.978, 0.982 and 0.993 respectively. The highest R^2 value between the actual and predicted values is 0.993, which belongs to the training test. In ANFIS-SUB, the best predictions were obtained by 12 fuzzy rules with a cluster radius of 0.4. Comparison of the ANFIS-FCM and ANFIS-SUB models showed that the ANFIS-FCM was slightly better than the ANFIS-SUB model with higher R^2 values, lower errors (RMSE, SSE) and rules. However, the results of comparison between the predicted and the actual data indicated that both models have a strong potential in estimating Cd in the groundwater with a high degree of accuracy. Sonmez et al. applied ANFIS to predict Cd concentrations in the Filyos River situated in Turkey. The result of the study approved a high correlation ($R^2 = 0.91$) between observed and predicted Cd concentrations. The outcomes revealed the reasonable estimates of the ANFIS model gave for the concentrations of Cd with a high degree of accuracy [10]. Ghadimi et al. used an ANN model to estimate heavy metals in the groundwater in Arak City, Iran. Their results showed that ANN is a reliable method for the prediction of heavy metals (HMs) in the groundwater with high accuracy [52]. In another study, Sari et al. used ANNs for predicting heavy metals

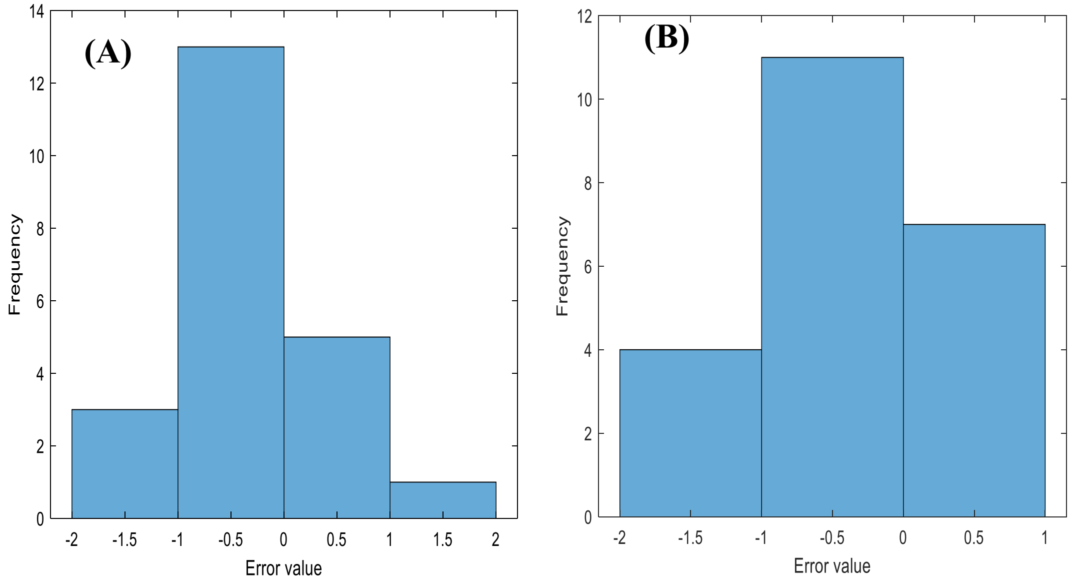

Fig. 11. ANFIS-SUB histogram of errors in original scale for total data(A), ANFIS-FCM histogram of errors for total data(B).

Table 3
Comparative performance of ANFIS-FCM and ANFIS-SUB models.

Indices	ANFIS-FCM Total	Train	Test	ANFIS-SUB Total	Train	Test
R^2	0.983	0.999	0.998	0.978	0.993	0.982
RMSE	0.669	0.068	1.56	0.702	0.43	1.43
SSE	9.846	0.085	9.74	10.85	3.34	8.14

concentration of soil samples obtained from different altitudes on Mount Ida. Result of the research showed that the computed relative errors were significantly low for each of the considered elements (Fe, Mn, and Zn); and error ranges were found to be 1.0–4.1%, 1.0–4.2%, 1.5–7.1%, respectively, for the training, testing, and holdout data [53]. Alizamir et al. have reported a good generalization performance of the ELM approach in surface water management the potential of the ANN-PSO model to predict the concentration of heavy metals in the Toyserkan Plain was useful to implement sustainable policies for groundwater management [54]. Fattahi et al. used three Modified Adaptive Neuro-Fuzzy Inference System (MANFIS) models MANFIS-GP, MANFIS-SCM, and MANFIS-FCM, for the estimation of metal concentrations in the Shur River. They result showed the superiority of the MANFIS-SCM model. Also, these results indicate that the MANFIS-SCM model has the potential for estimation of the metals with a high degree of accuracy and robustness [26]. Jafari et al. applied ANFIS and wavelet-ANFIS models based on FCM for predicting groundwater fluctuations. The maximum R^2 wasfound as 0.997 and 0.994 in the training and test stages and the best values of RMSE were 0.05 and 0.08 m, respectively. A comparison of the ANFIS and wavelet-ANFIS models showed the superiority of the latter model in modeling groundwater levels (GWL) because it employed the synergy of the FCM clustering technique and the wavelet transform [55]. Lu et al. applied the ANN and support vector machine (SVM) models to simulate HM concentration in an aquatic environment. The result of the study demonstrates that both ANN and SVM simulated concentrations of particulate HM well with Nash-Sutcliffe efficiency >0.8. Also, these models acted worse in simulating dissolved and total HM concentrations. Results proved that artificial intelligence-based models like ANN and SVM are good alternatives to simulate HM concentrations with limited monitoring data [56]. Comparing our ANFIS-FCM model with other studies, we found that our model exhibited a strong correlation ($R^2 = 0.982$) between the predicted values and actual experimental data, indicating high accuracy in predicting Cd levels. This finding is consistent with above mentioned studies that have used machine learning techniques to predict heavy metal concentrations in environmental samples, such as support vector regression (SVR) and artificial neural networks (ANNs). In terms of error, our model also demonstrated low sum square error (SSE) and root mean square error (RMSE), indicating good performance in predicting Cd concentrations. Overall, our study suggests that the ANFIS-FCM model is a promising approach for accurately predicting heavy metal concentrations in environmental samples.

5. Conclusions

Fuzzy clustering techniques allow the automatic generation of fuzzy models and can be utilized to predict water quality monitoring. In this study, Cd concentration was monitored in groundwater resources of Neyshabur city and a prediction was made using two different fuzzy models ANFIS-FCM and ANFIS-SUB. Three different performance criteria including the correlation coefficient (R^2), root mean square error (RMSE), and sum square error (SSE) were used for evaluating the performance of ANFIS-FCM and ANFIS-SUB models. The study demonstrated that the ANFIS-FCM model was effective in accurately predicting Cd levels, with a high degree of correlation ($R^2 = 0.982$) between the predicted values and actual experimental data. In general, the study discovered that using physicochemical parameters, the ANFIS-FCM model and ANFIS-SUB have a reliable ability to forecast the levels of Cd in groundwater. This suggests a hopeful technique for approximating Cd concentrations. Consequently, ANFIS-FCM and ANFIS-SUB are an easy way to produce valid results to assess the cadmium concentration of a groundwater area and to reduce time losses during the evaluation stage. In upcoming research, machine learning techniques can be utilized to forecast the levels of heavy metals in diverse environmental samples. Further future potential studies, (a) analyzing the impact of different input variables on the accuracy of adaptive fuzzy systems for modeling heavy metals in groundwater resources, may improve the results, (b) selection methods can be implemented to pinpoint the most significant factors that impact heavy metal concentrations, (c) comparing the performance of adaptive fuzzy systems with other machine learning techniques for modeling heavy metals in groundwater resources (d) Developing a real-time monitoring system based on adaptive fuzzy systems for detecting heavy metals contamination in groundwater resources.

Author contribution statement

Naghmeh Jafarzade: Conceived and designed the experiments; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Ozgur Kisi: Analyzed and interpreted the data; Wrote the paper.

Mahmood Yousefi: Performed the experiments; Wrote the paper.

Mansour Baziar: Analyzed and interpreted the data; Wrote the paper.

Vahide Oskoei: Performed the experiments; Wrote the paper.

Nilufar Marufi: Performed the experiments; Wrote the paper.

Ali Akbar Mohammadi: Conceived and designed the experiments; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Data availability statement

Data will be made available on request.

Additional information

No additional information is available for this paper.

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

The authors want to thank authorities of Neyshabur University of Medical Sciences for their comprehensives support for this study.

References

[1] M. Bigalke, et al., Accumulation of Cadmium and Uranium in Arable Soils in Switzerland, vol. 221, 2017, pp. 85–93.
[2] M. Shams, et al., Heavy metals exposure, carcinogenic and non-carcinogenic human health risks assessment of groundwater around mines in Joghatai, Iran 102 (8) (2022) 1884–1899.
[3] N. Jafarzadeh, et al., Non-carcinogenic risk assessment of exposure to heavy metals in underground water resources in Saraven, Iran: spatial distribution, monte-carlo simulation, sensitive analysis 204 (2022), 112002.
[4] A.A. Mohammadi, et al., Carcinogenic and Non-carcinogenic Health Risk Assessment of Heavy Metals in Drinking Water of Khorramabad, vol. 6, 2019, pp. 1642–1651. Iran.
[5] M. Esmaeilzadeh, et al., Investigation of the extent of contamination of heavy metals in agricultural soil using statistical analyses and contamination indices 25 (5) (2019) 1125–1136.
[6] V. Kumar, et al., Global evaluation of heavy metal content in surface water bodies: a meta-analysis using heavy metal pollution indices and multivariate statistical analyses, Chemosphere 236 (2019), 124364.
[7] M. Triassi, et al., Heavy metals in groundwater of southern Italy: occurrence and potential adverse effects on the environment and human health, Int. J. Environ. Res. Publ. Health 20 (3) (2023) 1693.
[8] K.L.M. De Jesus, et al., Neuro-Particle Swarm Optimization Based In-Situ Prediction Model for Heavy Metals Concentration in Groundwater and Surface Water 10 (2) (2022) 95.
[9] Z. Ullah, et al., Groundwater contamination through potentially harmful metals and its implications in groundwater management, Front. Environ. Sci. 10 (2022) 2077.
[10] A.Y. Sonmez, et al., An adaptive neuro-fuzzy inference system (ANFIS) to predict of cadmium (Cd) concentrations in the Filyos River, Turkey 18 (12) (2018) 1333–1343.
[11] M. Afzaal, et al., Heavy metals contamination in water, sediments and fish of freshwater ecosystems in Pakistan, Water Pract. Technol. 17 (5) (2022) 1253–1272.
[12] P. Li, et al., Sources and consequences of groundwater contamination, Arch. Environ. Contam. Toxicol. 80 (2021) 1–10.

[13] E. Shaji, et al., Arsenic contamination of groundwater: A global synopsis with focus on the Indian Peninsula 12 (3) (2021), 101079.
[14] D. Karunanidhi, et al., Chromium contamination in groundwater and Sobol sensitivity model based human health risk evaluation from leather tanning industrial region of South India, Environ. Res. 199 (2021), 111238.
[15] V. Goyal, et al., Appraisal of heavy metal pollution in the water resources of Western Uttar Pradesh, India and associated risks, Environmental Advances 8 (2022), 100230.
[16] A.A. Kana, Heavy metal assessment of groundwater quality in part of Karu, Central Nigeria, Water Pract. Technol. 17 (9) (2022) 1802–1817.
[17] A. Kubier, R.T. Wilkin, T.J.A.G. Pichler, Cadmium in Soils and Groundwater: a Review, vol. 108, 2019, 104388.
[18] B. Kiani, et al., Association between heavy metals and colon cancer: an ecological study based on geographical information systems in North-Eastern Iran, BMC Cancer 21 (1) (2021) 1–12.
[19] Q. Chen, et al., Hydroinformatics techniques in eco-environmental modelling and management 8 (4) (2006) 297–316.
[20] W.-B. Chen, W.-C.J.E.M. Liu, Assessment, Artificial neural network modeling of dissolved oxygen in reservoir 186 (2) (2014) 1203–1217.
[21] R. Haggerty, et al., Application of Machine Learning in Groundwater Quality Modeling-A Comprehensive Review, Water Research, 2023, 119745.
[22] F. Qaderi, E.J. J.o, C.P. Babanezhad, Prediction of the groundwater remediation costs for drinking use based on quality of water resource, using artificial neural network 161 (2017) 840–849.
[23] A. Mosavi, et al., Susceptibility prediction of groundwater hardness using ensemble machine learning models 12 (10) (2020) 2770.
[24] S. Moghanlo, et al., Using Artificial Neural Networks to Model the Impacts of Climate Change on Dust Phenomenon in the Zanjan Region, vol. 35, 2021, 100750 *north-west Iran*.
[25] S. Azimi, M.A. Moghaddam, S.H. J, J.o.c.h. Monfared, Prediction of annual drinking water quality reduction based on Groundwater Resource Index using the artificial neural network and fuzzy clustering 220 (2019) 6–17.
[26] H. Fattahi, et al., Multi-output adaptive neuro-fuzzy inference system for prediction of dissolved metal levels in acid rock drainage: a case study 6 (1) (2018) 121–132.
[27] X. Ding, et al., Comparison of Models for Spatial Distribution and Prediction of Cadmium in Subtropical Forest Soils, Guangdong, China 10 (9) (2021) 906.
[28] S. Fallahizadeh, et al., The effects of meteorological parameters on PM10: health impacts assessment using AirQ+ model and prediction by an artificial neural network (ANN), Urban Clim. 38 (2021), 100905.
[29] S. Mehdizadeh, B. Mohammadi, F. Ahmadi, Establishing coupled models for estimating daily dew point temperature using nature-inspired optimization algorithms, Hydrology 9 (1) (2022) 9.
[30] N. Patel, et al., Water table depth forecasting based on hybrid wavelet neural network model, in: Evolution in Computational Intelligence: Proceedings of the 9th International Conference on Frontiers in Intelligent Computing: Theory and Applications (FICTA 2021), Springer, 2022.
[31] S. Samantaray, A. Sahoo, D.P. Satapathy, Prediction of groundwater-level using novel SVM-ALO, SVM-FOA, and SVM-FFA algorithms at Purba-Medinipur, India, Arabian J. Geosci. 15 (8) (2022) 723.
[32] H. Tao, et al., Groundwater level prediction using machine learning models: a comprehensive review, Neurocomputing 489 (2022) 271–308.
[33] Y. Chen, et al., A review of the artificial neural network models for water quality prediction, Appl. Sci. 10 (17) (2020) 5776.
[34] H. Riahi-Madvar, et al., Short to long-term forecasting of river flows by heuristic optimization algorithms hybridized with ANFIS, Water Resour. Manag. 35 (2021) 1149–1166.
[35] U.K. Das, et al., Estimation of aquifer potential using BPNN, RBFN, RNN, and ANFIS, in: Smart Intelligent Computing and Applications: Proceedings of the Second International Conference on SCI, vol. 2, Springer, 2018, 2019.
[36] S. Emamgholizadeh, et al., Prediction of soil cation exchange capacity using enhanced machine learning approaches in the southern region of the Caspian Sea, Ain Shams Eng. J. 14 (2) (2023), 101876.
[37] F. Hadadi, R. Moazenzadeh, B. Mohammadi, Estimation of actual evapotranspiration: a novel hybrid method based on remote sensing and artificial intelligence, J. Hydrol. 609 (2022), 127774.
[38] A.M. Ahmed, S.M.A. Shah, Application of adaptive neuro-fuzzy inference system (ANFIS) to estimate the biochemical oxygen demand (BOD) of Surma River, Journal of King Saud University-Engineering Sciences 29 (3) (2017) 237–243.
[39] T.M. Akpomie, et al., Computer modelling of the concentration of heavy metals in artificial borings, World Journal of Analytical Chemistry 4 (1) (2016) 6–10.
[40] S. Areerachakul, Comparison of ANFIS and ANN for estimation of biochemical oxygen demand parameter in surface water, Int. J. Chem. Biomol. Eng. 6 (2012) 286–290.
[41] F.-J. Chang, et al., Assessment of arsenic concentration in stream water using neuro fuzzy networks with factor analysis, Sci. Total Environ. 494 (2014) 202–210.
[42] S. Dahiya, et al., Analysis of groundwater quality using fuzzy synthetic evaluation, J. Hazard Mater. 147 (3) (2007) 938–946.
[43] B.Y. Öztürk, et al., Fuzzy logic evaluation of heavy metal pollution of Apa Dam Lake, J. Appl. Biol. Sci. 8 (3) (2014) 14–20.
[44] M. Yousefi, et al., Data on water quality index for the groundwater in rural area Neyshabur County, Razavi province, Iran, Data Brief 15 (2017) 901–907.
[45] M. Rand, A.E. Greenberg, M.J. Taras, Standard Methods for the Examination of Water and Wastewater, 1976. Prepared and published jointly by American Public Health Association.
[46] J.-S. Jang, ANFIS: adaptive-network-based fuzzy inference system, IEEE transactions on systems, man, and cybernetics 23 (3) (1993) 665–685.
[47] P. Souza, G. Dotto, N. Salau, Artificial neural network (ANN) and adaptive neuro-fuzzy interference system (ANFIS) modelling for nickel adsorption onto agro-wastes and commercial activated carbon, J. Environ. Chem. Eng. 6 (6) (2018) 7152–7160.
[48] E.H. Ruspini, Numerical methods for fuzzy clustering, Inf. Sci. 2 (3) (1970) 319–350.
[49] M. Zarinbal, M.F. Zarandi, I. Turksen, Relative entropy fuzzy c-means clustering, Inf. Sci. 260 (2014) 74–97.
[50] A. Najah, et al., Performance of ANFIS versus MLP-NN dissolved oxygen prediction models in water quality monitoring, Environ. Sci. Pollut. Control Ser. 21 (2014) 1658–1670.
[51] M. Baziar, et al., Application of adaptive neural fuzzy inference system and fuzzy C-means algorithm in simulating the 4-chlorophenol elimination from aqueous solutions by persulfate/nano zero valent iron process, Eurasian J. Anal. Chem 13 (2018) 1–10.
[52] F. Ghadimi, Prediction of heavy metals contamination in the groundwater of Arak region using artificial neural network and multiple linear regression, J. Tethys 3 (3) (2015) 203–215.
[53] M. Sari, et al., Deciding heavy metal levels in soil based on various ecological information through artificial intelligence modeling, Appl. Artif. Intell. 36 (1) (2022), 2014189.
[54] M. Alizamir, S. Sobhanardakani, An artificial neural network-particle swarm optimization (ANN-PSO) approach to predict heavy metals contamination in groundwater resources, Jundishapur Journal of Health Sciences 10 (2) (2018).
[55] M.M. Jafari, et al., Application of a novel hybrid wavelet-ANFIS/fuzzy c-means clustering model to predict groundwater fluctuations, Atmosphere 12 (1) (2020) 9.
[56] H. Lu, et al., Simulating heavy metal concentrations in an aquatic environment using artificial intelligence models and physicochemical indexes, Sci. Total Environ. 694 (2019), 133591.