Water quality prediction using machine learning methods

Amir Hamzeh Haghiabi, Ali Heidar Nasrolahi and Abbas Parsaie

ABSTRACT

This study investigates the performance of artificial intelligence techniques including artificial neural network (ANN), group method of data handling (GMDH) and support vector machine (SVM) for predicting water quality components of Tireh River located in the southwest of Iran. To develop the ANN and SVM, different types of transfer and kernel functions were tested, respectively. Reviewing the results of ANN and SVM indicated that both models have suitable performance for predicting water quality components. During the process of development of ANN and SVM, it was found that tansig and RBF as transfer and kernel functions have the best performance among the tested functions. Comparison of outcomes of GMDH model with other applied models shows that although this model has acceptable performance for predicting the components of water quality, its accuracy is slightly less than ANN and SVM. The evaluation of the accuracy of the applied models according to the error indexes declared that SVM was the most accurate model. Examining the results of the models showed that all of them had some over-estimation properties. By evaluating the results of the models based on the DDR index, it was found that the lowest DDR value was related to the performance of the SVM model.

Key words | ANN, Dez catchment, GMDH, SVM, Tireh River

Amir Hamzeh Haghiabi
Ali Heidar Nasrolahi
Abbas Parsaie (corresponding author)
Water Engineering Department,
Lorestan University,
Khorramabad,
Iran
E-mail: *abbas_parsaie@yahoo.com*

INTRODUCTION

Water quality has a direct impact on public health and the environment. Water is used for various practices, such as drinking, agriculture, and industry. Recently, development of water sports and entertainment has greatly helped to attract tourists (Jennings 2007). Among various sources of water supply, due to easy access, rivers have been used more frequently for the development of human societies. Using other water resources such as groundwater and seawater sometimes assisted with problems. For example, using groundwater without suitable recharge will lead to land subsidence (Motagh *et al.* 2017) and using seawater is usually associated with pollution transmission (El-Kowrany *et al.* 2016). Therefore, the use of rivers has attracted attention. Several investigations related to rivers around the world have been conducted and a field of engineering named river engineering has been proposed. In river engineering, studies on morphological changes, sediment transport, water quality, and pollution transmission mechanisms are very important (Julien 2002; Dey 2014). Flow structure, sediment transport and morphology of rivers are investigated in the hydraulics of rivers in river engineering (Wu 2007). The study of water quality of rivers is a common theme in earth sciences. To evaluate the quality of rivers two approaches are considered, including measuring the water quality components and defining the mechanism of pollution transmission (Kashefipour 2002; Kashefipour & Falconer 2002; Naseri Maleki & Kashefipour 2012; Qishlaqi *et al.* 2016). Among water quality components, measuring the dissolved oxygen (DO), chemical oxygen demand (COD), biochemical oxygen demand (BOD), electrical conductivity (EC), pH, temperature, K, Na, Mg, etc. have been proposed (Şener *et al.* 2017). To this end, governments have constructed hydrometry stations along rivers that cross from urban areas, agro-industrial projects, industrial estates, and rivers that join dams' reservoirs (Herschy 1993; Kejiang 1993). In hydrometry stations, the water quality

doi: 10.2166/wqrj.2018.025

components are measured and the stage-discharge relation is defined. Obtained values from hydrometry stations contain basic information for feasibility studies and development of water conservation projects. Evaluation of water quality is a basic stage for development of agriculture projects in terms of determination of cropping pattern, type of irrigation system, and systems of water purification for industries (Chen *et al.* 2017). To investigate the mechanism of pollution transmission, in addition to field and laboratory experiments, advanced numerical methods such as computational hydraulic, image processing and GIS methods have been utilized (Parsaie & Haghiabi 2015, 2017a, 2017b). By reviewing the time history of water quality components, investigators have attempted to estimate future values. Nowadays, by advancing soft computing techniques in most areas of water and environmental engineering, researchers have attempted to accurately analyse time series of water quality components and their internal relation (May *et al.* 2008; Palani *et al.* 2008; Haghiabi 2016a, 2016b; Jaddi & Abdullah 2017). In this regard, Emamgholizadeh *et al.* (2013) used multilayer perceptron (MLP), radial basis network (RBF) and an adaptive neuro-fuzzy inference system (ANFIS) for water quality components of Karoon River. They stated that all applied models have suitable performance for prediction of water quality components; however, the MLP model was slightly more accurate. Shokoohi *et al.* (2017) managed the water quality of a water supply system. They considered this an optimization problem and used modern optimization methods to solve it. Zhang *et al.* (2010) introduced a new approach for water allocation. They considered water quality as one of the main factors in their approach. Nikoo & Mahjouri (2013) developed a Probabilistic Support Vector Machines (PSVMs) model associated with GIS technique for planning the classification and distribution of surface and groundwater water in Iran. They stated that the use of these two methods would provide accurate information for feasibility studies of water conservation projects. Heddam (2016a; 2016b; 2016c; 2016d; 2016e) utilized artificial neural networks for predicting the water quality components in several case studies. He stated that artificial intelligence techniques have suitable performance for modeling and predicting the internal relation between the water quality components and modeling their time series. Reviewing the literature shows that water quality assessment and prediction is an important factor for developing water conservation projects and, to this end, artificial intelligence techniques have been proposed. Hence, in this study the water quality components of Tireh River, one of the main rivers of Dez catchment (one of the major catchments in Iran), were predicted using a support vector machine, article neural network and group method of data handling.

MATERIAL AND METHODS

The aim of this study is the prediction of water quality components using artificial intelligence (AI) techniques including MLP, SVM, and group method of data handling (GMDH). Therefore, in the first part of this section, the studied area is introduced and then ranges of measured water quality components are presented. Overviews on applied AI models are then presented.

Case study and observed data

Dez catchment is one of the main watersheds of Iran. The catchment contains several intermittent and perennial streams. Tireh River is one of the main perennial streams of this catchment, located in Lorestan Province, which passes through the two major cities. Figure 1 shows the coordination of Tireh River. To determine the stage discharge relation and monitor the water quality components, the regional water authority (RWA) in Lorestan province (Iran) constructed hydrometry stations along this river. Constructed hydrometry stations by RWA are shown by triangular symbols. Measuring the stage discharge relation and water quality components by RWA was conducted monthly. It is noteworthy that in most months, several measurements have been recorded. The timeline of sampling is more than 55 years. The first measurement was reported in 1960 and to date, monitoring of this river is still ongoing. A summary of the measured water quality components by RWCA are given in Table 1. As given in the table, measurement parameters include temperature (T), pH, specific conductivity (EC), bicarbonate (HCO_3^{-1}), sulfates (SO_4^{-2}), chlorides (Cl), total dissolved solids (TDS), sodium (Na^{+}), magnesium (Mg^{+2}), calcium (Ca^{+2}).

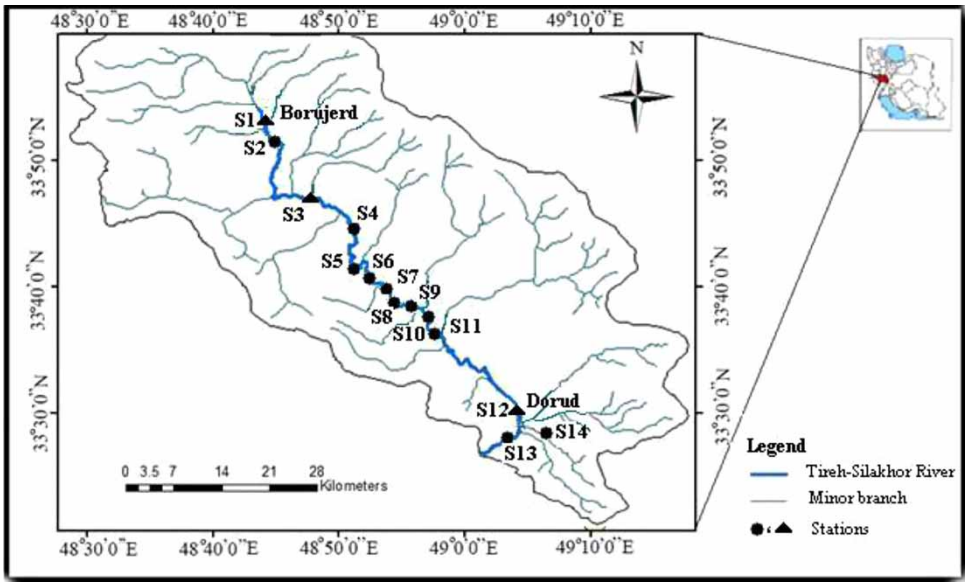

Figure 1 | Tireh River, its catchment boundaries and water quality sampling stations.

Table 1 | Summary of values of measured water quality parameters by RWA

Parameters	Min	Max	Mean	Variance
Q (m^3/s)	0.3	1,000.00	10.56	70.64
T.D.S (mg/L)	17.00	894.00	288.04	119.48
EC (μmohs/cm)	30.00	1,355.00	448.17	182.28
pH	6.20	8.85	7.80	0.41
HCO_3 (mg/L)	0.35	7.70	3.29	0.99
Cl (mg/L)	0.02	7.70	0.75	0.83
SO_4 (mg/L)	0.00	4.65	0.42	0.40
Ca (mg/L)	0.50	5.60	2.63	0.81
Mg (mg/L)	0.04	5.40	1.25	0.71
Na (mg/L)	0.01	4.00	0.58	0.67
SAR (mg/L)	0.01	3.02	0.40	0.43

Artificial neural network (ANN)

The initial development of the ANN used a data mining approach taken from the human brain's nervous system. This means that the structure of ANN consists of networks of neurons that work together in parallel (Zahiri *et al.* 2015). Mathematical operations of neurons are provided in Figure 2. As shown in this figure, inputs are multiplied by a series of weights and then summed by a constant value. Then a transfer function acts on the results of the previous

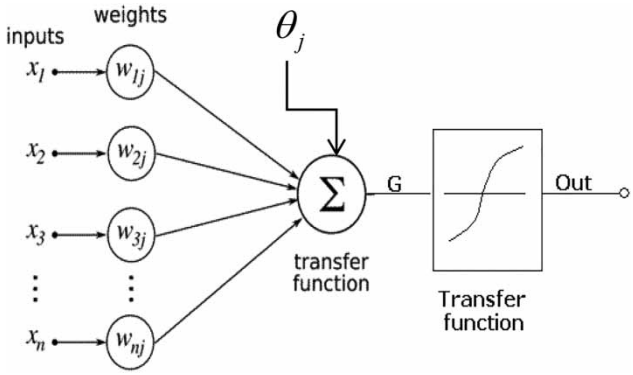

Figure 2 | A typical operation in a neuron in a neural network.

computation. The mathematical expression of neuron computation is given in Equation (1):

$$I_j = f\left(\sum_i w_{ij}\alpha_i + \theta_j\right) \quad (1)$$

As presented in Equation (1), α_i are inputs, w_{ij} are weights and θ_j are biases. f is a function that governs each cell and acts on its inputs. Most fame transfer functions are given in Equation (2). The structure of the neural network consists of three layers: the first layer, the middle layer and the output layer. The first layer is intended to introduce inputs. The middle layer may contain one or more

hidden layers. In other words, hidden layers are introduced as sub-layers in this area. The output layer is considered for summarization of the results of the parallel computation performed in the middle layer. The results of each neuron may be considered as inputs for the neurons existing in the next layer of a network or may be compared with the observed data. It is notable that values of weights and biases are justified via a training process (Haghiabi *et al.* 2016; Parsaie & Haghiabi 2017c).

$$\text{I.} \quad \text{Gaussian: } F(x) = a\,exp\left(-\frac{(x-b)^2}{c^2}\right)$$

$$\text{II.} \quad \text{Sigmoidal: } F(x) = \frac{1}{1+\exp(-x)}$$

$$\text{III.} \quad \text{Tansing: } F(x) = \frac{2}{(1+\exp(-2x))} - 1 \tag{2}$$

Support vector machine (SVM)

SVM is an artificial intelligence technique that has been widely utilized for pattern recognition, classification and regression. One of the well-known uses of SVM is function fitting, therefore, this part of SVM utilizing is called support vector regression (SVR). The goal of function fitting using SVM is minimizing the error (difference between the model output and observed data). This can be considered as an optimization problem and its mathematical expression is given as follows:

$$\begin{aligned} &\text{Minimize:} \quad R_{svm}(\omega, \xi^*) = \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \\ &\text{Subject} \\ &d_i - \omega\varphi(x_i) + b_i \le \varepsilon + \xi_i \\ &\omega\varphi(x_i) + b_i - d_i \le \varepsilon + \xi_i \\ &\xi_i, \xi_i^* \ge 0, \quad i = 1, \ldots, l, \end{aligned} \tag{3}$$

where ω is a normal vector, $(1/2)\|\omega\|^2$ is the regularization factor, C is the error penalty factor, b is a bias, ε is the error function, x_i is the input vector, d_i is the target value, l is the number of elements in the training data set, $\varphi(x_i)$ is a feature space, and ξ_i and $\xi * i$ are upper and lower excess deviation (Azamathulla & Wu 2011; Azamathulla *et al.* 2014). Most well-known kernel functions are given as follows:

I. Linear kernel: $K(x_i, x_j) = x_i^T x_j$
II. Polynomial kernel: $K(x_i, x_j) = (x_i^T x_j + \gamma)^d, \quad \gamma > 0$
III. RBF kernel: $K(x_i, x_j) = exp(-\gamma\|x_i - x_j\|)^2, \quad \gamma > 0$
IV. Sigmoid kernel: $k(x_i, x_j) = \tanh(\gamma x_i^T x_j + r), \quad \gamma > 0$

where variables x_i and x_j are inputs, and γ is the regularization parameter. Lagrange multipliers are presented as $\alpha_i = \alpha_i - \alpha_i^*$. The accuracy of prediction is based on the selection of three parameters, that is, γ, ε and C, whose values are determined using a firefly algorithm (Noori *et al.* 2011; Azamathulla 2013; Dehghani *et al.* 2014; Noori *et al.* 2015; Azamathulla *et al.* 2016; Gocic *et al.* 2016).

Group method of data handling (GMDH)

The GMDH method was introduced by Ivakhnenko (1971). The GMDH tries to estimate complex systems with a network of simple operations. The GMDH is a feed-forward neural network. In the GMDH model each pair of inputs are introduced to a neuron. This approach, the opposite of ANN models where all inputs are entranced in neurons, significantly reduces the computational cost. Figure 3 shows a schema of a GMDH-type neural network. The governing equation on each neuron is a quadratic polynomial and its mathematical expression is given in Equation (4):

$$\bar{y} = G(x_i, x_j) = w_0 + w_1 x_i + w_2 x_j + w_3 x_i^2 + w_4 x_j^2 + w_5 x_i x_j \tag{4}$$

The coefficients of equation are justified using the least square (LS) method. The LS method is given below. As mentioned previously, all the data related to the pairs of inputs are passed through Equation (4). Therefore, all the processes of LS operation can be rewritten as below:

$$y \approx \bar{y} = [w_0\, w_1\, w_2\, w_3\, w_4\, w_5] \begin{bmatrix} 1 \\ x_i \\ x_j \\ x_i^2 \\ x_j^2 \\ x_i x_j \end{bmatrix} \Rightarrow y = w^T x_{ij} \tag{5}$$

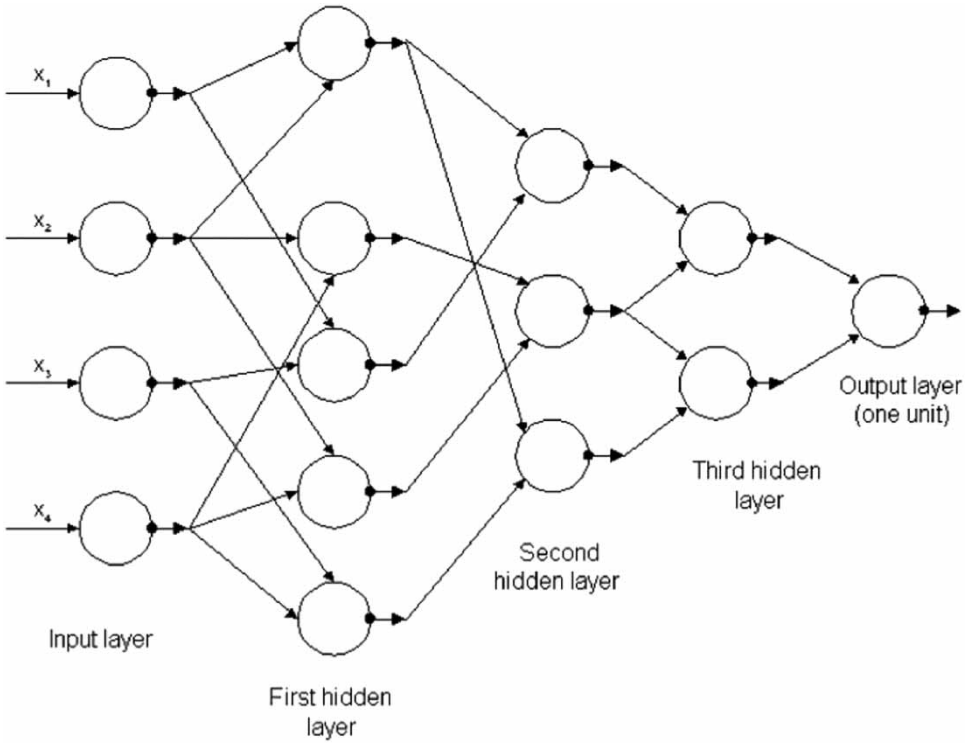

Figure 3 | A GMDH-type neural network.

Table 2 | Summary of performance of applied models for prediction of water quality component

		GMDH			SVM			ANN		
Row	**Output**	**Stage**	**R^2**	**RMSE**	**K-F**	**R^2**	**RMSE**	**T-F**	**R^2**	**RMSE**
1	Ca	Train	0.83	0.333	RBF	0.96	0.160	Tansig	0.92	0.238
		Test	0.85	0.313		0.94	0.193		0.84	0.295
2	Cl	Train	0.87	0.297	RBF	0.97	0.147	Tansig	0.94	0.193
		Test	0.88	0.312		0.95	0.210		0.96	0.178
3	E_C	Train	0.97	29.71	RBF	0.97	26.35	Tansig	0.98	23.88
		Test	0.99	13.84		0.98	28.81		0.96	35.14
4	HCO_3	Train	0.87	0.355	RBF	0.96	0.190	Tansig	0.96	0.192
		Test	0.89	0.334		0.95	0.219		0.92	0.290
5	Mg	Train	0.76	0.345	RBF	0.93	0.184	Tansig	0.92	0.197
		Test	0.79	0.332		0.93	0.199		0.90	0.212
6	Na	Train	0.78	0.317	RBF	0.94	0.157	Tansig	0.85	0.250
		Test	0.77	0.326		0.93	0.186		0.86	0.251
7	SO_4	Train	0.36	0.321	RBF	0.77	0.190	Tansig	0.74	0.185
		Test	0.26	0.297		0.68	0.204		0.68	0.240
8	TDS	Train	0.97	19.34	RBF	0.98	16.40	Tansig	0.98	16.60
		Test	0.99	7.22		0.97	22.31		0.97	19.39
9	pH	Train	0.25	0.358	RBF	0.59	0.270	Tansig	0.29	0.33
		Test	0.22	0.361		0.33	0.340		0.29	0.355

$$\left.\begin{array}{l} y_1 = w^T X_1 \\ y_2 = w^T X_2 \\ \quad \vdots \\ y_n = w^T X_n \end{array}\right\} \Rightarrow [y_1, y_2, \ldots, y_n] = w^T [X_1, X_2, \ldots, X_n]$$

$$\Rightarrow Y = w^T X \tag{6}$$

$$\overbrace{Y}^{\text{Targets}} = \underbrace{w^T}_{\text{Weighted coefficients}} \overbrace{X}^{\text{Regressors}}$$

$$YX^T = w^T XX^T$$

$$YX^T (XX^T)^{-1} = w^T \underbrace{XX^T (XX^T)^{-1}}_{I} \Rightarrow w^T = Y \underbrace{X^T (XX^T)^{-1}}_{X^\dagger} \tag{7}$$

To prepare the GMDH, a script was written in the Matlab and 'pinv' was the main built-in function that was used for solving $X^\dagger$. For developing the GMDH model, in the first layer pairs of inputs are selected and then all of them are introduced in neurons, individually. For the next layer, only neurons that have adequacy criteria are selected. The adequacy criteria is defined based on the error of neurons, in other words, each neuron that has less error is selected to develop the next layer (Zahiri *et al.* 2014; Parsaie *et al.* 2016, Parsaie & Haghiabi 2017c).

Development of models

The first stage of development of AI models is the preparation of the dataset. In this stage, the collected dataset should be divided into two groups – training and testing. The training and testing dataset are used for the calibration and validation of applied models, respectively. Depending on the simulation conditions regarding time series modeling or function fitting, the approach of assigning a dataset to each group are different. In time series modeling, the history of collecting data should be considered and shuffling the dataset is not correct, whereas for function fitting (identifying the internal relations of components to each other) using data shuffling idea is allowed. Usually for both scenarios, about 70–80% of the dataset is assigned for calibration and the remaining 20–30% for validation. The next step for developing the AI models, such as ANN and SVM, is designing the architecture of the network. The structure of such models includes the number of hidden layers, the number of neurons per layer, and the type of transfer function.

RESULTS AND DISCUSSION

In this part of the paper, the results of prediction of the internal relations between the water quality components are presented. To develop an optimal model, an approach that was introduced by Parsaie & Haghiabi (2017c) was considered. They stated that for developing the ANN, some steps should be considered to reduce the trial and error process. They stated that for the initial design of ANN model, after dataset division, in the first step one hidden layer consisting of numbers of neurons equal to input features is considered. At this stage, the performance of different

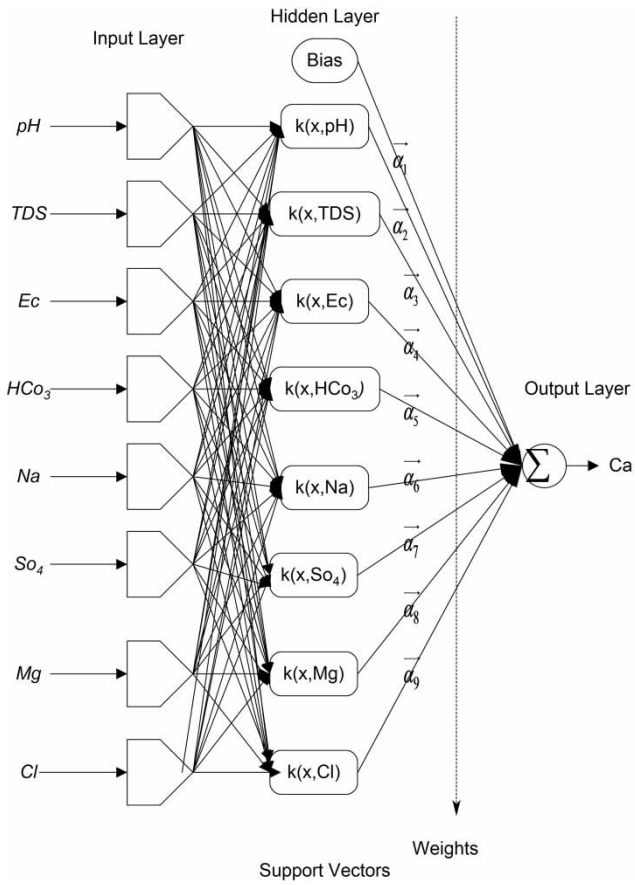

Figure 4 | Structure of SVM model for predicting water quality component (Ca).

transfer functions is evaluated and the best ones are chosen. In the next step, the size of the network is modified to improve the precision of the developed model. To this end, the numbers of neurons or number of hidden layers would increase. The last two stages of this approach are also applicable to the design of SVM. The main point relating to the design of the SVM is defining the kernel function. In this study, 80% of the dataset was used for training and the remaining 20% for testing. These two groups of datasets were used for developing the ANN, SVM and GMDH. For developing the GMDH model as stated in the Material and methods section, the designer only controls the threshold values and develops the network structure. The RMSE index was chosen for threshold values. The RMSE values of GMDH given in Table 2 in the testing stage were chosen as the threshold criteria. A summary of results of each applied model for predicting the water quality components are given in Table 2. For example, as presented in this table, for predicting the Ca, other water quality components including Cl, EC, HCO_3, Mg, Na, SO_4, TDS, pH were considered as inputs. Table 2 shows that for predicting the Ca, the SVM is most accurate in comparison with others (GMDH and ANN). As presented in Table 2, the best performance of ANN with coefficient of determination (0.92 and 0.84) and root means square error (0.238 and 0.295) in training and testing stages is related to the tansig function as best transfer function. It is notable that the structure of ANN, after the trial and error process, consisted of two hidden layers where its first and second hidden layers included eight and three neurons, respectively. A comparison of three models in terms of predicting the Cl declared that the SVM model has the best performance. For predicting the Ec, all three models have suitable performance. For predicting the HCO_3, the performances of SVM and ANN are close together and their precision is greater than the GMDH. This result was repeated for Mg. For predicting the Na, the SVM has the best performance. For estimation of SO_4, the best performance was related

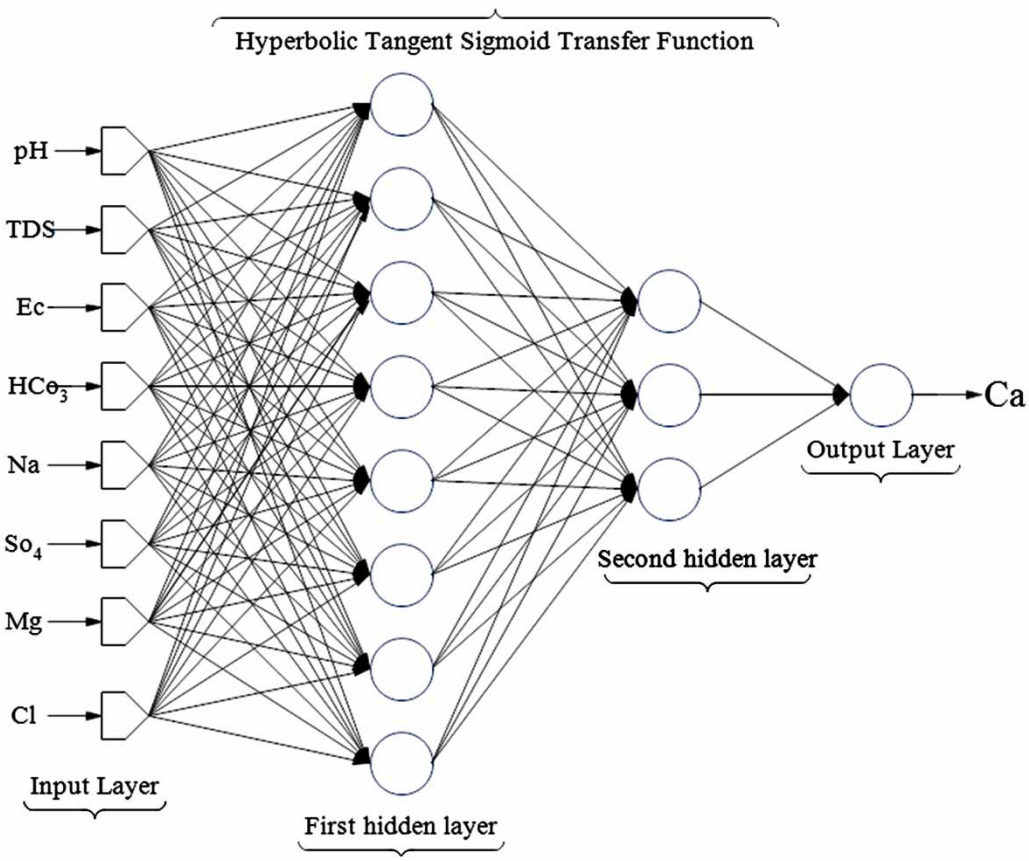

Figure 5 | Structure of ANN model for predicting water quality component (Ca).

SVM. For estimation of TDS, all three models have suitable performance and their accuracy is close together. For estimation of pH, the best performance is related to SVM. The accuracy of this model in the prediction of pH indicates that its accuracy has been reduced by a small amount. The structure of SVM and ANN models for predicting the Ca are shown in Figures 4 and 5, respectively. Reviewing Table 2 indicates that the RBF and tansing functions have the best performance in comparison with other tested kernel and transfer functions. The results of three applied AI models for estimation of Ca and Cl in training and testing stages are shown in Figure 6. To present further information related to the performance of applied models throughout the dataset, the DDR index, introduced by Noori *et al.* (2017), was calculated. This index is calculated using Equation (8). This index shows the performances of applied models related to properties of lower and over estimation. Results of DDR for testing stages of applied models for all water quality components are shown in Figure 7. As shown in this figure, the most amount of data dispersion related DDR index is related to the ANN model and the lowest data dispersion is related to the SVM. Although the accuracy of GMDH was less than SVM and somewhere less than the ANN, the data dispersion related DDR values are close to SVM. Reviewing Figure 7 shows that all three models have a slight over-estimation property:

$$DDR = \left(\frac{\text{Predicted Value}}{\text{Observed Value}}\right) - 1 \tag{8}$$

CONCLUSIONS

In this paper, the performance of artificial intelligence techniques including GMDH, SVM and ANN were evaluated to

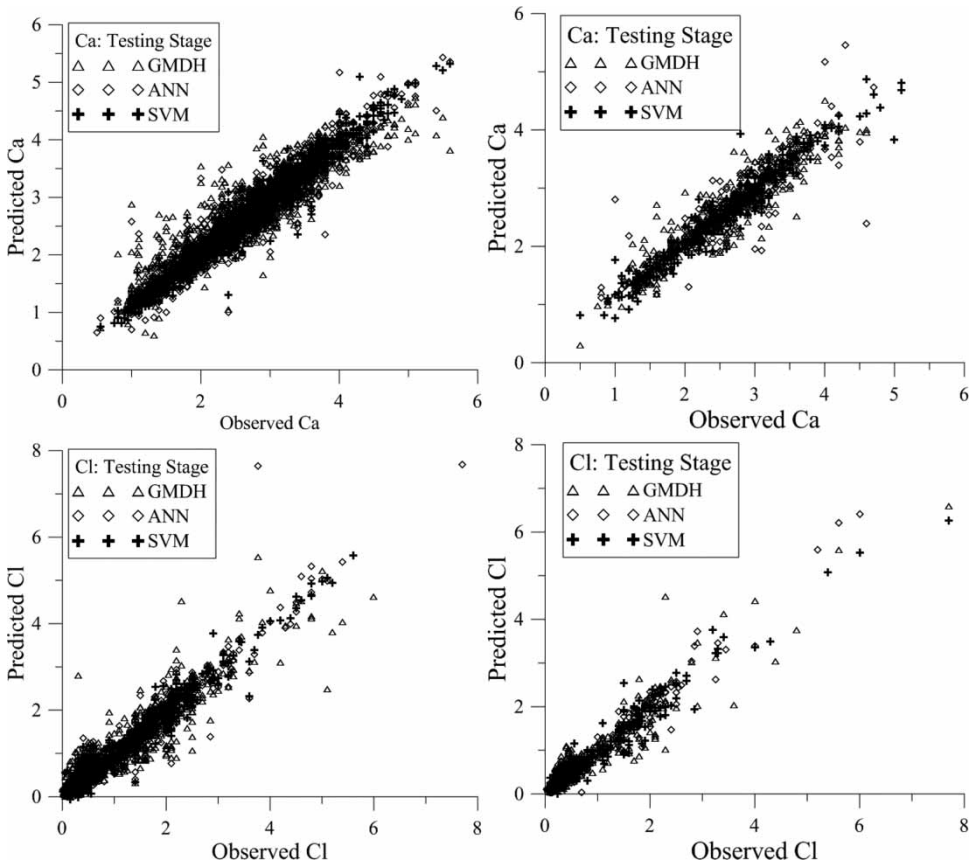

Figure 6 | Results of applied AI models versus observed data (Ca and Cl).

predict the water quality components of Tireh River (Iran). To this end most dataset related well-known components, such as pH, SO_4, Na, Ca, Cl, Mg, HCO_3 etc., were collected. Results indicated that the applied models have suitable performance for predicting water quality components, however, the best performance was related to the SVM. Reviewing the

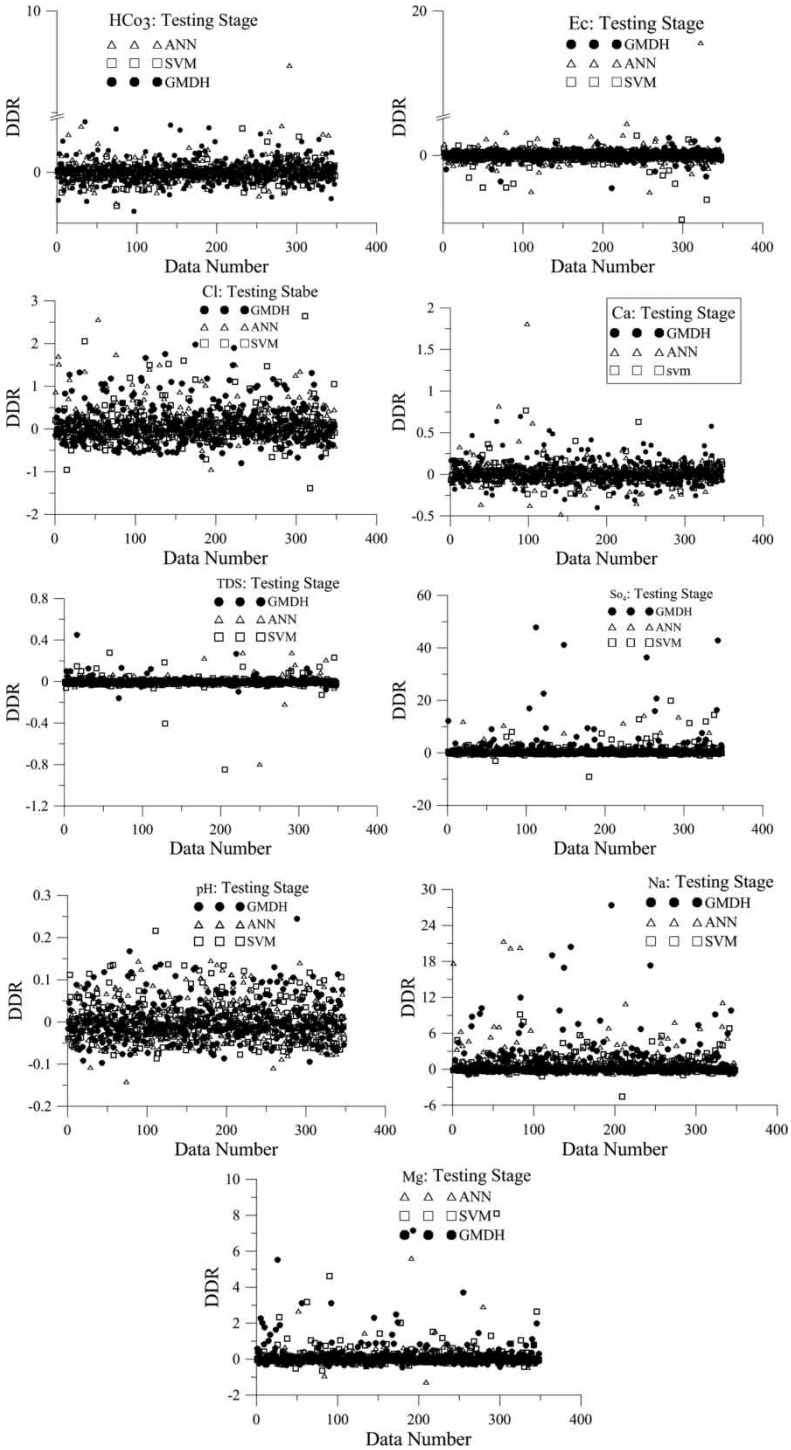

Figure 7 | Results of DDR index for applied AI models for predicting the water quality component.

structure of SVM showed that the best accuracy was related to the RBD as kernel function. Results of ANN indicated that its accuracy is acceptable for practical purposes. The best performance of tested transfer function was related to tansig. The lowest accuracy of models was related to GMDH. The DDR index of results of applied models shows that all three models slightly over-estimate. Comparison of the performance of GMDH, SVM and ANN according to DDR shows that the data dispersion of SVM was less than the others. Although the accuracy of model GMDH is less than that of model SVM, their DDR indices are close together. Furthermore, comparison of the performance of applied models indicated that the outcomes of GMDH and SVM models were more reliable in comparison with ANN.

REFERENCES

Azamathulla, H. M. 2013 2 – A Review on Application of Soft Computing Methods in Water Resources Engineering A2 – Yang, Xin-She. In: *Metaheuristics in Water, Geotechnical and Transport Engineering* (A. H. Gandomi, S. Talatahari & A. H. Alavi, eds). Elsevier, Oxford, pp. 27–41.

Azamathulla, H. M. & Wu, F.-C. 2011 Support vector machine approach for longitudinal dispersion coefficients in natural streams. *Appl. Soft Comput.* **11** (2), 2902–2905.

Azamathulla, H. M., Yusoff, M. A. M. & Hasan, Z. A. 2014 Scour below submerged skewed pipeline. *J. Hydrol.* **509**, 615–620.

Azamathulla, H. M., Haghiabi, A. H. & Parsaie, A. 2016 Prediction of side weir discharge coefficient by support vector machine technique. *Water Sci. Technol. Water Supply* **16** (4), 1002–1016.

Chen, X., Chen, Y., Shimizu, T., Niu, J., Nakagami, K. i., Qian, X., Jia, B., Nakajima, J., Han, J. & Li, J. 2017 Water resources management in the urban agglomeration of the Lake Biwa region, Japan: an ecosystem services-based sustainability assessment. *Sci. Total Environ.* **586** (Suppl. C), 174–187.

Dehghani, M., Saghafian, B., Nasiri Saleh, F., Farokhnia, A. & Noori, R. 2014 Uncertainty analysis of streamflow drought forecast using artificial neural networks and Monte-Carlo simulation. *Int. J. Climatol.* **34** (4), 1169–1180.

Dey, S. 2014 *Fluvial Hydrodynamics: Hydrodynamic and Sediment Transport Phenomena.* Springer, Berlin Heidelberg.

El-Kowrany, S. I., El- Zamarany, E. A., El-Nouby, K. A., El-Mehy, D. A., Abo Ali, E. A., Othman, A. A., Salah, W. & El-Ebiary, A. A. 2016 Water pollution in the Middle Nile Delta, Egypt: an environmental study. *J. Adv. Res.* **7** (5), 781–794.

Emamgholizadeh, S., Kashi, H., Marofpoor, I. & Zalaghi, E. 2013 Prediction of water quality parameters of Karoon River (Iran) by artificial intelligence-based models. *Int. J. Environ. Sci. Technol.* **11** (3), 645–656.

Gocic, M., Shamshirband, S., Razak, Z., Petković, D., Ch, S. & Trajkovic, S. 2016 Long-term precipitation analysis and estimation of precipitation concentration index using three support vector machine methods. *Adv. Meteorol.* doi:10.1155/2016/7912357.

Haghiabi, A. H. 2016a Modeling river mixing mechanism using data driven model. *Water Resour. Manage.* **31** (3), 811–824.

Haghiabi, A. H. 2016b Prediction of longitudinal dispersion coefficient using multivariate adaptive regression splines. *J. Earth Syst. Sci.* **125** (5), 985–995.

Haghiabi, A. H., Azamathulla, H. M. & Parsaie, A. 2016 Prediction of head loss on cascade weir using ANN and SVM. *ISH J. Hydr. Eng.* **23** (1), 102–110.

Heddam, S. 2016a Generalized regression neural network based approach as a new tool for predicting total dissolved gas (TDG) downstream of spillways of dams: a case study of Columbia River Basin Dams, USA. *Environ. Process.* **4** (1), 235–253.

Heddam, S. 2016b Multilayer perceptron neural network-based approach for modeling phycocyanin pigment concentrations: case study from lower Charles River buoy, USA. *Environ. Sci. Pollut. Res.* **23** (17), 17210–17225.

Heddam, S. 2016c New modelling strategy based on radial basis function neural network (RBFNN) for predicting dissolved oxygen concentration using the components of the Gregorian calendar as inputs: case study of Clackamas River, Oregon, USA. *Model. Earth Syst. Environ.* **2** (4), 162–167.

Heddam, S. 2016d Secchi disk depth estimation from water quality parameters: artificial neural network versus multiple linear regression models? *Environ. Process.* **3** (2), 525–536.

Heddam, S. 2016e Simultaneous modelling and forecasting of hourly dissolved oxygen concentration (DO) using radial basis function neural network (RBFNN) based approach: a case study from the Klamath River, Oregon, USA. *Model. Earth Syst. Environ.* **2** (3), 117–135.

Herschy, R. 1993 National and international standards in streamflow measurement. *Flow Meas. Instrum.* **4** (1), 53–55.

Ivakhnenko, A. G. 1971 Polynomial theory of complex systems. *IEEE Trans. Syst. Man Cybernet.* **1** (4), 364–378.

Jaddi, N. S. & Abdullah, S. 2017 A cooperative-competitive master-slave global-best harmony search for ANN optimization and water-quality prediction. *Appl. Soft Comput.* **51**, 209–224.

Jennings, G. 2007 *Water-based Tourism, Sport, Leisure, and Recreation Experiences.* Elsevier, Oxford.

Julien, P. Y. 2002 *River Mechanics.* Cambridge University Press, Cambridge, p. 435.

Kashefipour, S. M. 2002 *Modelling Flow, Water Quality and Sediment Transport Processes in Reverine Basins.* PhD thesis, Cardiff University, Cardiff.

Kashefipour, S. M. & Falconer, R. A. 2002 Longitudinal dispersion coefficients in natural channels. *Water Res.* **36** (6), 1596–1608.

Kejiang, C. 1993 Flow measurement in large rivers in China. *Flow Meas. Instrum.* **4** (1), 47–50.

May, R. J., Dandy, G. C., Maier, H. R. & Nixon, J. B. 2008 Application of partial mutual information variable selection to ANN forecasting of water quality in water distribution systems. *Environ. Model. Softw.* **23** (10–11), 1289–1299.

Motagh, M., Shamshiri, R., Haghshenas Haghighi, M., Wetzel, H.-U., Akbari, B., Nahavandchi, H., Roessner, S. & Arabi, S. 2017 Quantifying groundwater exploitation induced subsidence in the Rafsanjan plain, southeastern Iran, using InSAR time-series and in situ measurements. *Eng. Geol.* **218**, 134–151.

Naseri Maleki, M. & Kashefipour, S. M. 2012 Application of numerical modeling for solution of flow equations and estimation of water quality pollutants in rivers (Case study: Karkheh River). *Civil Environ. Eng.* **42.3** (68), 51–60.

Nikoo, M. R. & Mahjouri, N. 2013 Water quality zoning using probabilistic support vector machines and self-organizing maps. *Water Resour. Manage.* **27** (7), 2577–2594.

Noori, R., Karbassi, A. R., Moghaddamnia, A., Han, D., Zokaei-Ashtiani, M. H., Farokhnia, A. & Gousheh, M. G. 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** (3–4), 177–189.

Noori, R., Yeh, H.-D., Abbasi, M., Kachoosangi, F. T. & Moazami, S. 2015 Uncertainty analysis of support vector machine for online prediction of five-day biochemical oxygen demand. *J. Hydrol.* **527**, 833–843.

Noori, R., Ghiasi, B., Sheikhian, H. & Adamowski, J. F. 2017 Estimation of the dispersion coefficient in natural rivers using a granular computing model. *J. Hydraul. Eng.* **143** (5). doi:10.1061/(ASCE)HY.1943-7900.0001276.

Palani, S., Liong, S.-Y. & Tkalich, P. 2008 An ANN application for water quality forecasting. *Mar. Pollut. Bull.* **56** (9), 1586–1597.

Parsaie, A. & Haghiabi, A. 2015 The effect of predicting discharge coefficient by neural network on increasing the numerical modeling accuracy of flow over side weir. *Water Resour. Manage.* **29** (4), 973–985.

Parsaie, A. & Haghiabi, A. H. 2017a Computational modeling of pollution transmission in rivers. *Appl. Water Sci.* **7** (3), 1213–1222.

Parsaie, A. & Haghiabi, A. H. 2017b Numerical routing of tracer concentrations in rivers with stagnant zones. *Water Sci. Technol. Water Supply* **17** (3), 825–834.

Parsaie, A. & Haghiabi, A. H. 2017c Improving modelling of discharge coefficient of triangular labyrinth lateral weirs Using SVM, GMDH and MARS techniques. *Irrigation and Drainage* **66** (4), 636–654.

Parsaie, A., Haghiabi, A. H., Saneie, M. & Torabi, H. 2016 Applications of soft computing techniques for prediction of energy dissipation on stepped spillways. *Neural Comput. Appl.* doi.org/10.1007/s00521-00016-02667-z.

Parsaie, A., Azamathulla, H. M. & Haghiabi, A. H. 2017a Physical and numerical modeling of performance of detention dams. *J. Hydrol.* doi.org/10.1016/j.jhydrol.2017.01.018.

Parsaie, A., Ememgholizadeh, S., Haghiabi, A. H. & Moradinejad, A. 2017b Investigation of trap efficiency of retention dams. *Water Sci. Technol. Water Supply.* doi: 10.2166/ws.2017.109.

Qishlaqi, A., Kordian, S. & Parsaie, A. 2016 Hydrochemical evaluation of river water quality – a case study. *Appl. Water Sci.* **7** (5), 2337–2342.

Şener, Ş., Şener, E. & Davraz, A. 2017 Evaluation of water quality using water quality index (WQI) method and GIS in Aksu River (SW-Turkey). *Sci. Total Environ.* **584–585**, 131–144.

Shokoohi, M., Tabesh, M., Nazif, S. & Dini, M. 2017 Water quality based multi-objective optimal design of water distribution systems. *Water Resour. Manage.* **31** (1), 93–108.

Wu, W. 2007 *Computational River Dynamics.* CRC Press, London, UK.

Zahiri, A., Azamathulla, H. M. & Ghorbani, K. 2014 Prediction of local scour depth downstream of bed sills using soft computing models. In: *Computational Intelligence Techniques in Earth and Environmental Sciences* (T. Islam, P. K. Srivastava, M. Gupta, X. Zhu & S. Mukherjee, eds). Springer Netherlands, Dordrecht, Switzerland, pp. 197–208.

Zahiri, A., Dehghani, A. A. & Azamathulla, H. M. 2015 Application of gene-expression programming in hydraulic engineering. In: *Handbook of Genetic Programming Applications* (A. H. Gandomi, A. H. Alavai & C. Ryan, eds). Springer International Publishing, Switzerland, pp. 71–97.

Zhang, W., Wang, Y., Peng, H., Li, Y., Tang, J. & Wu, K. B. 2010 A coupled water quantity–quality model for water allocation analysis. *Water Resour. Manage.* **24** (3), 485–511.

First received 25 July 2017; accepted in revised form 21 December 2017. Available online 19 January 2018