RESEARCH

Qanat discharge prediction using a comparative analysis of machine learning methods

Saeideh Samani[1] · Meysam Vadiati[2] · Ozgur Kisi[3,4] · Leyla Ghasemi[5] · Reza Farajzadeh[6]

Abstract

The Qanat (also known as kariz) is one of the significant water resources in many arid and semiarid regions. The present research aims to use machine learning techniques for Qanat discharge (QD) prediction and find a practical model that predicts QD well. Gene expression programming (GEP), artificial neural network (ANN), group method of data handling (GMDH), least-square support vector machine (LSSVM) and adaptive neuro-fuzzy inference system (ANFIS), are employed to predict one-, two-, and five-months time-step ahead QD in an unconfined aquifer. QD for one, two, and three lag-times (QD_{t-1}, QD_{t-2}, QD_{t-3}), QD for adjacent Qanat, the main meteorological components (T_t, ET_t, P_t) and GWL for one, two, and three lag-times are utilized as input dataset to accomplish accurate QD prediction. The GMDH model, according to its best results, had promising accuracy in predicting multi-step ahead monthly QD, followed by the LSSVM, ANFIS, ANN and GEP, respectively.

Keywords Arid and semiarid · Groundwater hydraulics · Prediction · Qanat discharge · Machine learning

Introduction

Nations in arid and semiarid regions have dealt with the severe water scarcity crisis and water supply procedures. One such reliable and natural water resources supplier is groundwater. Ancient Iranians invented Qanat as a water supplier for drinking and agricultural purposes in cities and villages. Qanat system was successively employed mainly in the Middle East, North Africa, southern Spain, and other arid and semiarid regions. Qanats provide groundwater without energy consumption and extract water by gravity (Azari Rad et al. 2018; Naghibi et al. 2018; Samani et al. 2023a; Sedghi and Zhan 2024). More than economic benefits, Qanats are groundwater regulators which dewater the groundwater level in some areas with high groundwater levels. Compared to water well drilling, Qanats have fewer limitations in shallow groundwater and aquifers with low hydraulic conductivity. Qanats are dried due to extreme climate change and groundwater overexploitation (Sedghi and Zhan 2020). Despite their importance in the water cycle and people's life in the middle east, little scientific research has explored the hydrological and hydrogeological conditions of Qanats and, more specifically, modeling methods to predict Qanat discharge (QD); which was the pivot motivation for the current study.

In standard groundwater research, machine learning methods are promising tools for quantitative evaluation and predicting hydrogeological conditions and groundwater level fluctuations (Cui et al. 2022; Dehghani and Torabi Poudeh 2022). For instance, the main goal of modeling could be considering the Qanat discharge prediction based on different hydrologic scenarios and stresses and

✉ Saeideh Samani
saeideh.samani@dal.ca

1. Department of Civil and Resource Engineering, Dalhousie University - Sexton Campus, 1360 Barrington Street, Halifax, NS B3H 4R2, Canada

2. Hubert H. Humphrey Fellowship Program, Global Affairs, University of California, Davis, 10 College Park, Davis, CA 95616, USA

3. Department of Civil Engineering, University of Applied Sciences, 23562 Lübeck, Germany

4. Department of Civil Engineering, Ilia State University, N, 32 Ilia Chavchavadze Avenue, Tbilisi State, Georgia

5. University of Applied Science, Leeuwarden, Netherlands

6. Kharazmi university, Tehran, Iran

Springer

evaluating the hydrological effects of nearby Qanats on each other (Naghibi et al. 2018). The non-availability or insufficient data could be the most significant limitation of the mathematical models, particularly the comprehensive complex dataset to simulate the anthropical factors. Based on the prior studies, machine learning methods could provide a considerable advantage in modeling the pattern of the inputs-outputs datasets (Antonopoulos and Gianniou 2022; Pham et al. 2022; Poursaeid et al. 2022; Sun et al. 2022). The nonlinear time-variant behavior of hydrogeological systems is complicated enough to be solved with standard statistical methods (Najafabadipour et al. 2022). Then, researchers have tried to use various machine learning methods such as artificial neural networks (ANNs), gene expression programming (GEP), group method of data handling (GMDH), adaptive neuro-fuzzy inference system (ANFIS), least-square support vector machine (LSSVM) for groundwater level (GWL) prediction (Samani et al. 2013; Bahmani and Ouarda 2021; Ghazi et al. 2021a; Kiyani et al. 2022; Mozaffari et al. 2022; Poursaeid et al. 2022; Tao et al. 2022). The overall results of the prior studies revealed the ability of the machine learning models to cope with the complicated behavior of the hydrogeological systems and provide satisfactory results with the less-required dataset.

The ANN is a significant machine learning (ML) model that handles difficult and complex hydrological issues and problems based on statistical algorithms. ANN is the most applied and practical model for predicting groundwater's quantitative characteristics and conditions due to the architecture and codes' simplicity and availability (Iqbal et al. 2020; Banadkooki et al. 2020; Ahmadi et al. 2022; Samani et al. 2022). Based on the literature review, the studies focused on GWL modeling using ANFIS methods have also increased. The ANFIS model could provide simulated time series with high accuracy and precision for hydrological time series with different time steps, i.e., monthly, daily, or weekly (Moravej et al. 2020; Sridharam et al. 2021; Samantaray et al. 2022; Tao et al. 2022). The GMDH model is a promising tool for finding nonlinear relationships of hydrological systems. GMDH has been extensively applied in various areas of hydrological studies, such as river flow prediction, management, and soil and sediment (Lin et al. 2020; Khodakhah et al. 2022; Jaafari et al. 2022; Mulashani et al. 2022; Nadiri et al. 2022). In addition, various studies have been performed on groundwater level prediction exploiting the GMDH approach (Moghaddam et al. 2021; Arya Azar et al. 2022; Tao et al. 2022; Samani et al. 2023b). LSSVM has been implemented for predicting GWL and reported that this method improves the accuracy compared to ANN in predicting GWL (Miraki et al. 2019; Guzman et al. 2019; Khedri et al. 2020).

Several works have been reported in the literature regarding the application of ANN, ANFIS, GMDH and LSSVM in hydrogeological issues, civil engineering, soil science and water quality (Tayebi et al. 2019; Lin et al. 2020; Tao et al. 2022; Samantaray and Sahoo 2023; Samantaray et al. 2023 a and b; Samani 2024; Tao et al. 2024). GEP model has been widely used in groundwater modeling (Ghazi et al. 2021b) and in other fields such as leak detection of water distribution networks (Tijani and Zayed 2022), modeling soil enzymes (Ebrahimi et al. 2021), simulation of the subgouge soil deformation in the sand (Azimi and Shiri 2020), and prediction of daily dew point temperature (Mehdizadeh et al. 2017).
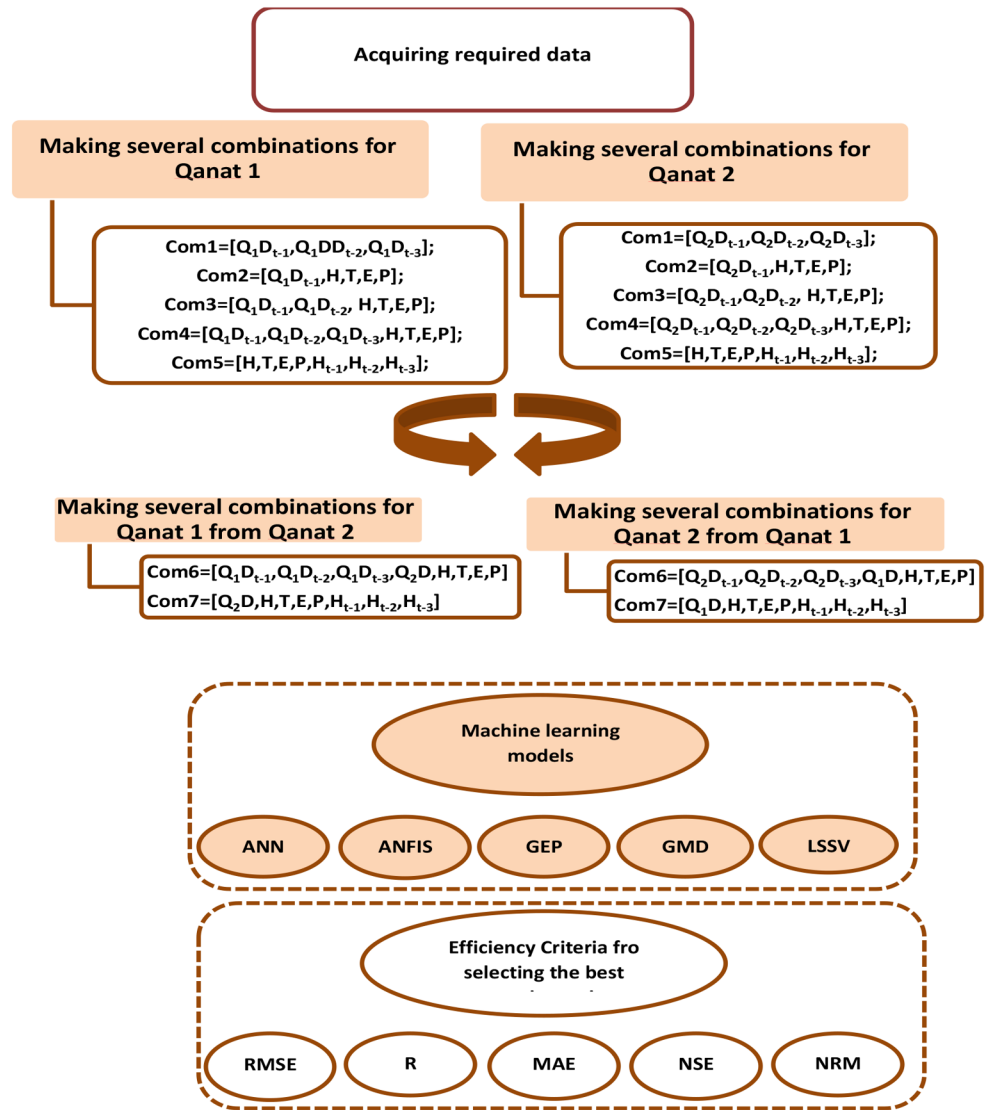
Despite the critical role of Qanats in providing water in arid and semi-arid regions, scientific research has not extensively explored the hydrological and hydrogeological dynamics specific to Qanats. This gap is significant given the pressing challenges posed by climate change and groundwater overexploitation, which threaten the sustainability of these ancient systems. Furthermore, while machine learning methods have been successfully applied to predict groundwater levels in various contexts, their application to Qanat discharge prediction remains limited. This study aims to bridge these gaps by (1) Evaluating the effectiveness of various machine learning models in predicting Qanat discharge, which has been less studied compared to other groundwater systems; (2) Identifying the most influential hydrological and hydrogeological parameters affecting Qanat discharge, an area not thoroughly investigated in previous Qanat studies; (3) Demonstrating the interplay between Qanats and adjacent hydrological features, which has not been sufficiently modeled or understood in the context of Qanat sustainability and efficiency. By focusing on these research gaps, this study seeks to advance the understanding of Qanat systems and improve the predictive capabilities of hydrological models in arid environments, ultimately contributing to better water resource management in these vulnerable regions. Figure 1 illustrates the procedural outline of the implemented ML methods.

Methods

The selection of machine learning methods for predicting Qanat discharge was based on several criteria aimed at capturing the complex, non-linear relationships inherent in hydrological data. The chosen methods are known for their robustness in handling various types of input data and their ability to model complex systems effectively. Here are the specific reasons for selecting each method:

Artificial Neural Network (ANN): ANN was chosen due to its proven ability in numerous hydrological studies to

Fig. 1 The methodological procedure of the QD models

Making several combinations for Qanat 1

Com1=$[Q_1D_{t-1},Q_1DD_{t-2},Q_1D_{t-3}]$;
Com2=$[Q_1D_{t-1},H,T,E,P]$;
Com3=$[Q_1D_{t-1},Q_1D_{t-2},H,T,E,P]$;
Com4=$[Q_1D_{t-1},Q_1D_{t-2},Q_1D_{t-3},H,T,E,P]$;
Com5=$[H,T,E,P,H_{t-1},H_{t-2},H_{t-3}]$;

Making several combinations for Qanat 2

Com1=$[Q_2D_{t-1},Q_2D_{t-2},Q_2D_{t-3}]$;
Com2=$[Q_2D_{t-1},H,T,E,P]$;
Com3=$[Q_2D_{t-1},Q_2D_{t-2},H,T,E,P]$;
Com4=$[Q_2D_{t-1},Q_2D_{t-2},Q_2D_{t-3},H,T,E,P]$;
Com5=$[H,T,E,P,H_{t-1},H_{t-2},H_{t-3}]$;

Making several combinations for Qanat 1 from Qanat 2

Com6=$[Q_1D_{t-1},Q_1D_{t-2},Q_1D_{t-3},Q_2D,H,T,E,P]$
Com7=$[Q_2D,H,T,E,P,H_{t-1},H_{t-2},H_{t-3}]$

Making several combinations for Qanat 2 from Qanat 1

Com6=$[Q_2D_{t-1},Q_2D_{t-2},Q_2D_{t-3},Q_1D,H,T,E,P]$
Com7=$[Q_1D,H,T,E,P,H_{t-1},H_{t-2},H_{t-3}]$

Machine learning models

ANN   ANFIS   GEP   GMD   LSSV

Efficiency Criteria fro selecting the best

RMSE   R   MAE   NSE   NRM

model complex and non-linear relationships between inputs and outputs. ANNs are flexible and can be trained to learn patterns from historical data, making them suitable for predicting Qanat discharge which is influenced by a variety of hydrological variables.

Adaptive Neuro-Fuzzy Inference System (ANFIS): ANFIS combines the learning capabilities of neural networks with the linguistic rule presentation of fuzzy logic. This hybrid approach allows it to effectively handle uncertainty and model the nonlinear relationships between hydrological parameters. Its success in previous studies involving groundwater level prediction under uncertain conditions justified its selection.

Gene Expression Programming (GEP): GEP was selected for its ability to create models that can evolve over time, allowing for better adaptation to changing patterns in data. This feature is particularly useful in the context of Qanats,

where hydrological inputs may change due to environmental factors or human activities.

Group Method of Data Handling (GMDH): GMDH is known for its self-organizing capabilities, making it suitable for modeling complex systems where the relationships between parameters are not fully understood. It was chosen for its ability to generate a model with optimal complexity, reducing the risk of overfitting which is crucial in accurately predicting Qanat discharge.

Least Square Support Vector Machine (LSSVM): LSSVM was included for its robustness in dealing with small datasets and its effectiveness in solving non-linear regression and classification problems. Given the limited data availability for Qanats, LSSVM's capability to provide high accuracy with fewer data points is highly advantageous.

These methods were also chosen because of their complementary strengths, providing a comprehensive approach to the modeling process. By comparing their performance,

we aim to identify the most effective model(s) for QD prediction, contributing to more reliable water resource management strategies.

Artificial neural network (ANN)

The ANN model, which is adapted based on brain behavior, performs in two stages, (i) obtaining knowledge from the environment via a learning procedure and (ii) using the inter-neuron link to gather the obtained knowledge (Haykin 2004; Patel et al. 2022; Sreelakshmi and Shaji 2022). The whole process consists of five phases: choosing the input dataset, choosing a proper framework, neural network designing, the process of training and testing, and lastly, model evaluation (Sahoo and Jha 2013). As the well-accepted ANN model in simulating hydrological phenomena, multilayer perceptron (MLP) was applied in the present work (McGarry et al. 1999). The MLP model consists of three main layers: input, hidden (middle) and output. The quantity of layers and neurons is crucial to obtaining an optimal model framework. In the present study, one hidden layer was utilized since it could be adequate for QD prediction based on earlier similar studies. The Levenberg-Marquardt (LM) algorithm was applied for the MLP training. MATLAB$^{®}$ (Mathworks 2014) software package was used to create ML models in the present study. The general structures/procedures of the developed models are given in Fig. 2a-e.Adaptive Neuro Fuzzy Inference System (ANFIS).

The ANFIS, as a particular ML method, could take advantage of the ANN and the fuzzy inference system (Jang 1993). The ANFIS is a flexible statistical method that could recognize complex nonlinear and uncertain patterns considering ambiguity between variables without completely knowing the nature of the problem. The Sugeno system is used in the ANFIS, and generally, the ANFIS structure is composed of 5-layers (Fig. 2b): fuzzy membership, fuzzification, normalization, defuzzification, and output (Jang 1993; Wee et al. 2021). The subtractive clustering (SC) is utilized in the present study to split the input dataset dimension into n-divided specific areas by assessing the n-dimensional input dataset to create different specific clusters. The cluster radius could be selected from 0 to 1 to optimize the cluster centroid's influence range. Recognizing the optimum cluster radius is vital to identifying the clusters' number, which was obtained by the trial-error procedure in the current paper.

Gene expression programming (GEP)

Koza (1992) proposed Genetic Programming (GP) as a generalization of the Genetic Algorithm (GA) (Goldberg 1989), which utilizes a "parse tree" form to search in the solution space, and GEP uses the benefits of GA and GP.

The GEP model was first introduced by Ferreira (2001), which is an intelligent evolutionary model. The initial phase in the GEP algorithm is the initial population generation that could be achieved randomly or with some data about the problem. Next, the chromosomes are expressed to create the tree expression. Then, the model results are assessed based on a fitness function to define the degree of satisfaction. If the reasonable model outcomes (desired threshold of fitness function or generation) are found, the evolution procedure will be interrupted, and the best-achieved results at this step are reported. However, if the stop provisions are not found, the best result for the current model generation will be maintained. The modeling procedure is repeated for several generations until the solution is found (Ferreira 2001). The structure, input and target variable, and position function of the GEP model differ from other ML methods, such as ANN. The optimum structure of the GEP model and coefficients are defined during the training procedure. The nature of the GEP provides further flexibility to the model. The overall structure of the GEP method is given in Fig. 2c.

Group method of data handling (GMDH)

Ivakhnenko (1968) introduced the GMDH method, which employs a self-organizing model (SOM) to address intricate and nonlinear problems, specifically tackling prediction, classification, and various other challenges. The main inputs, the quantity of hidden layers and neurons, and network structure are essentially specified in the GMDH method. However, the GMDH as a polynomial neural network resembles ANN models. Mueller et al. (1998) declared that statistical analysis and ANN are deductive methods that could not uncover complex objects because they need a lot of a priori information. Instead, the GMDH model as a regression-based method could combine the advantages of both methods (Lemke 1997). Therefore, GMDH could cover the deficits of ANN, while statistical neural networks could slightly solve them, and all model structures in the GMDH model could be specified by default. Nariman-Zadeh et al. (2002) provided more detailed information about the GMDH model. The framework of the GMDH model is presented in Fig. 2d.

Least square support vector machine (LSSVM)

Vapnik (1998) introduced the theory and basic concepts of the SVM. The SVM's general outline is better than the ANN since it is based on structural risk minimization, whereas the ANN employs experimental risk minimization. The SVM's main process consists of the selection of support vectors supporting the model structure and identifying the weights. A comprehensive mathematical framework of the SVM was
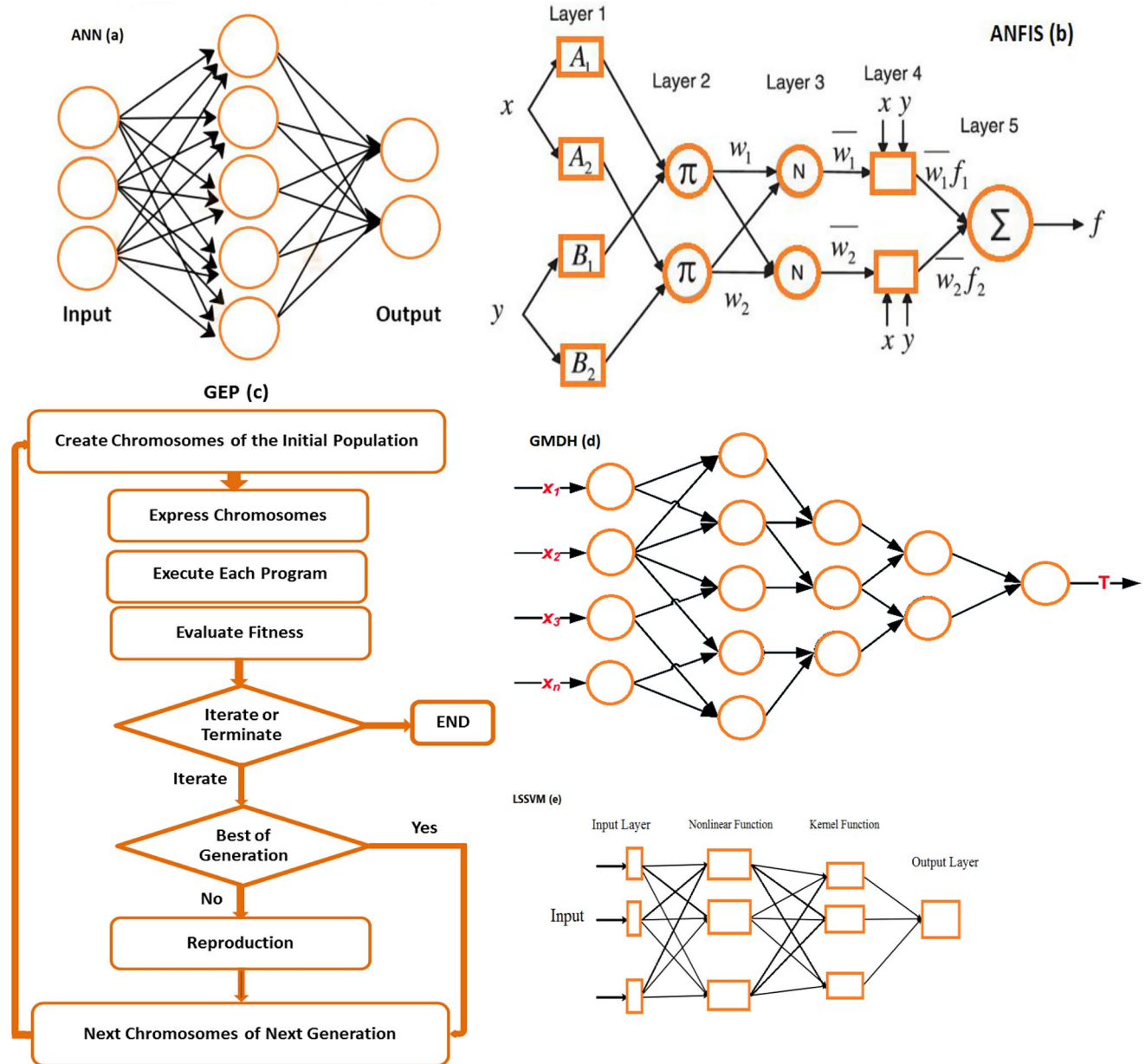
Fig. 2 The general structure of ML models: ANN (**a**), ANFIS (**b**), GEP(**c**), GMDH (**d**) and LSSVM (**e**)

suggested by Vapnik (1998). Then, Suykens and Vande-walle (1999) proposed the LSSVM based on the SVM. LSSVM is a robust method to resolve function estimation, nonlinear classification, and density estimation of hydrological problems. LSSVM solves linear programming subjects by modifying inequality constraints in the SVM to equality constraints (Kumar and Kar 2009). Also, the LSSVM is faster than the SVM (Fels and Ghorfi 2022; Gu et al. 2010).

Different algorithms were proposed to solve the dual optimization issue of SVM models. The Sequential Minimal Optimization (SMO) is the latest learning algorithm for SVM, which employs an analytical phase (Platt 1999).

SMO can directly solve the SVM problem without utilizing a quadratic optimizer or additional matrix space employed in this work. The result of LSSVM differs strongly on the appropriate selection of the kernel function and modifying the appropriate parameters of C and γ. The polynomial kernel function used in this study for LSSVM since its outstanding outcomes in QD prediction is dependent on the study area's dataset. Also, the optimum parameters of the LSSVM model are achieved based on a trial-and-error process (Suryanarayana et al. 2014). The LSSVM was applied using codes of the LIBSVM library introduced by Chang

and Lin (2011). The overall structure of the LSSVM model is given in Fig. 2e.

Qanat system

The Qanat system denotes a conventional reliance on the water supply and agricultural usage and is a remarkable typical aspect of Iran's climate and landscape. However, in many areas, the Qanat is now quickly substituted by deep wells.

Qanats have been built by the hand of experienced laborers with extensive knowledge of geology and engineering. As shown in Fig. 3, a Qanat system comprises the following four main elements:

1. Mother well: a dug deep into the groundwater level. Qanats conduct groundwater by gravity, and the mother well is typically built on sedimentary formation at the baseline of mountains. The most profound mother well, about 300 m, belongs to a 2,700-year-old Qanat in Gonabad in Iran (Boustani 2008).
2. Outlet: the area where water emerges to the surface. There are regularly various nominee locations for the groundwater outlet. The last position is determined concerning several factors, such as the vicinity to the positions of water utilization and the gallery's slope.
3. Gallery: Once the outlet and mother well's location are determined, a slightly sloped tunnel is started to build from the outlet in the direction of the mother well. The selection of the gallery's slope could be a trade-off between erosion and sedimentation, which usually slope of Qanats is about 0.5%. The gallery length differs from a hundred meters to kilometers, and the most extended gallery of Qanats is about 120 km, belongs to Zarach in Iran (Molle et al. 2004) and,
4. Shafts: are a string of vertical wells constructed along the Qanat's gallery between the outlet and the mother well at a 20–50 m distance to assist sediment removal
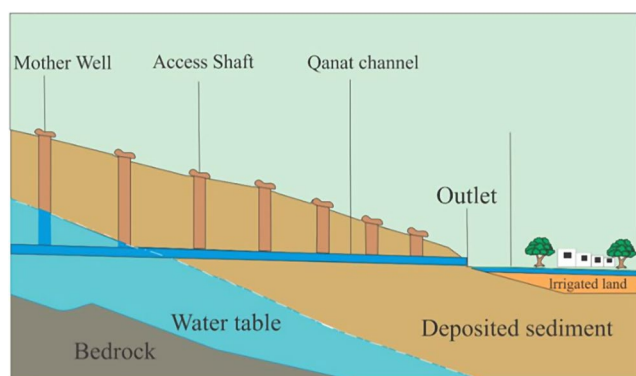

Fig. 3 Cross section of typical Qanat in an unconfined aquifer

and support air circulation and entry for laborers while constructing the gallery and after finishing the construction to monitor and maintain the Qanat.To simplify the subject, Qanat could be defined as a practically horizontal tunnel with some shaft wells which transmit groundwater from the aquifer to the surface. Qanat water leaks into the tunnel in the saturated zone and then streams down the tunnel to the outlet.

Also, Qanats can be considered a drainage system that drains groundwater to supply agricultural and drinking usage (Boustani 2008; Ganjeizadeh Rohani et al. 2024; Mohajerani et al. 2024; Nasiri and Mafakheri 2015; Yazdi and Khaneiki 2016). A cross-section of a typical Qanat in an unconfined aquifer and two intersecting Qanats near the town of Meybod, Yazd Province, Iran is given in Fig. 3. It should be mentioned that in a Qanats, the discharge is a function of specific storage of the aquifer, the length of the gallery and groundwater level; subsequently groundwater fluctuations can directly affect the Qanats discharge. Although the Qanat has provided water for Iranian for over two thousand years, the difficulty of controllability of water discharge has supposed that Qanat is now unsuitable for the optimum usage of water resources.

Study area

The study area of Razan-Ghahavand is located in Hamedan province, with an area of 3084 square kilometers. Water requirements for residents of this area are supplied by 1788 wells, 104 springs, and 96 Qanats. The groundwater system is experiencing significant tension due to an imbalance between extraction and infiltration, resulting in excessive exploitation of groundwater resources. A detailed location of the study area is given in Fig. 4. The elevation of the study area differs from 1581 to 2741 m above mean sea level, and the mean annual temperature and precipitation of the study region, respectively are 10° C and 245 mm.

Data and preprocessing

This study utilizes time series data collected from the Razan-Ghahavand Aquifer over a period of 18 years, from 2003 to 2021. The data, provided by the Hamedan Regional Water Authority, includes monthly measurements of Qanat discharge (QD), groundwater levels (GWL), and meteorological parameters such as temperature (T), precipitation (P), and evapotranspiration (ET). Each selected variable plays a specific role in the hydrological dynamics of Qanats. Discharge for adjacent Qanat (QD1 and QD2): Reflects the
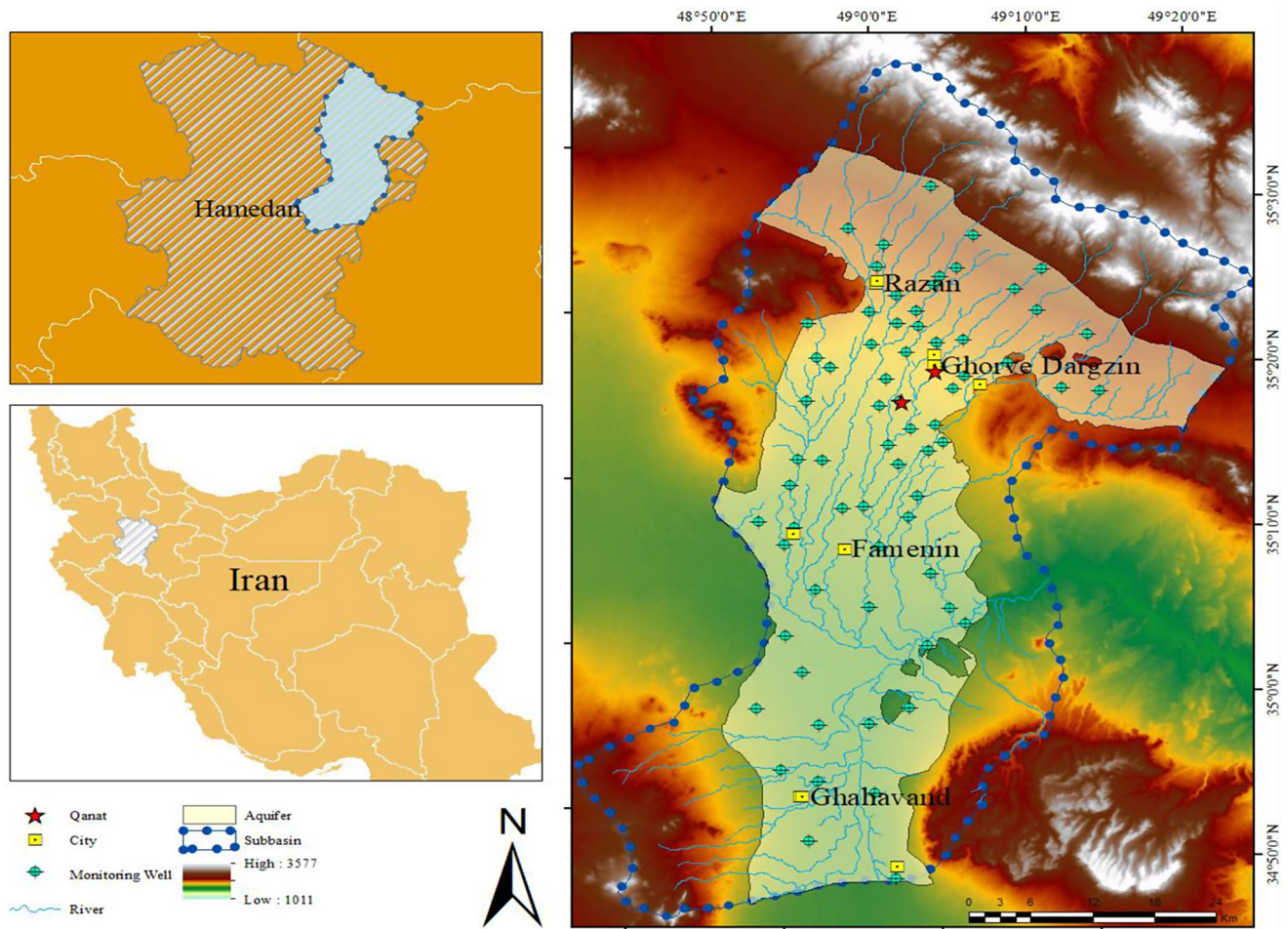
Fig. 4 The geographical map of the study area

effect of neighboring discharged on the each Qanat; Ground-water Level (GWL): Indicates the aquifer's state, influencing QD through hydraulic connectivity; Meteorological Data: Temperature, precipitation, and evapotranspiration directly affect evaporation and recharge rates, hence influencing QD. The integration of these variables suggests their interplay determines the aquifer's response to natural and anthropogenic changes, making them critical for accurate QD predictions. The output variable for models is the Qanat discharge (QD), predicted at one, two, and five-month lead times.

For accurate modeling of Qanat discharge (QD) using groundwater levels (GWL) and hydrological time series data, the optimal number of lagged inputs that significantly influence the current values must be determined. The Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) analyses were employed to identify these effective lags. The correlation between time series observations at different time lags is measured by ACF, which indicates how past values are related to future values without distinguishing between direct and indirect

effects. Conversely, PACF is used to measure the correlation between observations at different lags while controlling for the values of the intervening observations, thereby isolating the direct effects. Significant autocorrelations up to lag 3 were demonstrated in the ACF and PACF plots for both Qanat 1 and Qanat 2, as well as for GWL. Based on this observation, the inclusion of lags 1, 2, and 3 of QD and GWL as predictors in the machine learning models was decided.

Model development

Monthly local QD (QD_{t-1}, QD_{t-2}, QD_{t-3}) and GWL for one, two and three lag-times (GWL_{t-1}, GWL_{t-2}, GWL_{t-3}), temperature, (T), precipitation (P), evapotranspiration (ET) and QD of neighboring Qanat were considered as inputs for one-, two-, and five-month ahead QD predictions for two adjacent Qanats in Razan-Ghahavand Aquifer. This study explored GWL fluctuations based on a representative hydrograph. The Hamedan Regional Water Authority published

the Qanat discharge and monthly GWLs for 18 years from 2003 to 2021. To evaluate the models' efficiency in predicting QD, the input-output time series was split into two sections, 70% for training and 30% for the testing phase.

Model implementation

Selecting the most appropriate combination of input datasets is an essential step in ML modeling.

Considering unrelated datasets with time lags as input datasets in the scenarios reduces the results' reliability and complicates the models' structure. The general correlation analysis shows that QD in both Q_1 and Q_2 Qanats in the Razan-Ghahavand strongly correlated with time series data of QD of neighboring Qanat (Q_1 and Q_2) and GWL fluctuations. Moreover, meteorological datasets were additionally used for QD prediction.

Different input combinations were evaluated using the predictive input parameters with different lag time steps from one-month "QD_{t-1}" to three-month prior "QD_{t-3}" to predict QD with various lead times (one- to five-month ahead) (Table 1).

Results and discussion

This research focuses on utilizing machine learning methods to accurately predict QD and identify the most effective model for practical application. The study employs several machine learning techniques, including GEP, ANN, GMDH, LSSVM, ANFIS. These methods are used to forecast QD one, two, and five months in advance within an unconfined aquifer.

Designing an optimum structure for each model is an important phase of the modeling process since an inappropriate structure of the model could trigger over/ under-fitting issues. In the present study, a three-layered ANN model was deemed for QD prediction; Initial outcomes revealed that one hidden layer was sufficient to get a relationship between the predictor inputs and QD. Generally, a trial-and-error procedure was applied to identify the number of neurons in the hidden layer. This information is provided in the third column of Tables 2, 3, 4, 5, 6 and 7 for Qanat 1 and 2. Optimizing the cluster radius is a substantial concern for the efficient ANFIS structure. Smaller radii make numerous small clusters and, subsequently, many rules, whereas large radii cause a few large clusters to get fewer rules (Sanikhani and Kisi 2012). The optimal LSSVM parameters were determined as C= [0.2, 1] and $\gamma = 5$ by trial-and-error procedure. RBF is chosen as the appropriate kernel function.

Model comparison according to computational effort and Run Times

Computational cost is regularly a substantial constraint of prediction models. In the present study, we use different ML models to predict QD to achieve this goal; the computation times for the ANN, ANFIS, GEP, GMDH, and LSSVM models for the first combination and one month ahead, respectively, were recorded as 4.70, 3.90, 6.17, 1.32, and 0.48 s. The outcomes reveal that the LSSVM is faster compared to other implemented ML models. Moreover, the appropriate iteration (epoch) number, an important factor for ML models, could improve the model accuracy in both training and validation stages and could prevent overtraining. Adjusting various models with various forms and it is apparent that 100–200 iterations are enough for calibration of all developed models in predicting QD in both Qanats. Th epoch number for all developed ANN models was less than 100.

Comparison of the implemented models

This study aims to consider the application of various machine learning models ANN, ANFIS, GEP, and LSSVM to predict QD levels up to five months beyond data records. In the first part, various input combinations were examined by the applied models.

Table 1 Different combinations for both models

Combination	Input data		Group
	Q_1	Q_2	
Combination 1	$[Q_1D_{t-1}, Q_1DD_{t-2}, Q_1D_{t-3}]$	$Com1=[Q_2D_{t-1}, Q_2D_{t-2}, Q_2D_{t-3}]$	1
Combination 2	$[Q_1D_{t-1}, GWL, T,E, P]$	$Com2=[Q_2D_{t-1}, GWL, T,E, P]$	2
Combination 3	$[Q_1D_{t-1}, Q_1D_{t-2}, GWL, T,E, P]$	$Com3=[Q_2D_{t-1}, Q_2D_{t-2}, GWL, T,E, P]$	2
Combination 4	$[Q_1D_{t-1}, Q_1D_{t-2}, Q_1D_{t-3}, GWL, T,E, P]$	$Com4=[Q_2D_{t-1}, Q_2D_{t-2}, Q_2D_{t-3}, GWL, T,E, P]$	2
Combination 5	$[GWL, T,E, P, GWL_{t-1}, GWL_{t-2}, GWL_{t-3}]$	$Com5=[GWL, T,E, P, GWL_{t-1}, GWL_{t-2}, GWL_{t-3}]$	3
Combination 6	$[Q_1D_{t-1}, Q_1D_{t-2}, Q_1D_{t-3}, Q_2D, GWL, T,E, P]$	$Com6=[Q_2D_{t-1}, Q_2D_{t-2}, Q_2D_{t-3}, Q_1D, GWL, T,E, P]$	
Combination 7	$[Q_2D, GWL, T,E, P, GWL_{t-1}, GWL_{t-2}, GWL_{t-3}]$	$Com7=[Q_1D, GWL, T,E, P, GWL_{t-1}, GWL_{t-2}, GWL_{t-3}]$	3

Table 2 The results of the different combinations in predicting one- ahead QD for Qanat 1

Qanat 1	Input	Properties	Lt+1 Training					Test				
			R	RMSE	NRMSE	MAE	NS	R	RMSE	NRMSE	MAE	NS
ANN	Combination 1	2	0.71	9.5	0.13	7.37	0.5	0.72	3.57	0.19	2.81	0.46
	Combination 2	3	0.85	7.27	0.1	5.25	0.71	0.77	3.56	0.19	2.72	0.47
	Combination 3	4	0.85	7.07	0.09	5.3	0.72	0.84	2.69	0.14	2.14	0.69
	Combination 4	5	0.87	6.97	0.09	4.85	0.73	0.63	4.44	0.23	3.62	0.17
	Combination 5	6	0.84	7.97	0.11	6.38	0.65	0.65	3.79	0.2	2.98	0.4
	Combination 6	6	0.88	6.55	0.09	4.84	0.76	0.71	4.03	0.21	3.55	0.32
	Combination 7	0.5	0.85	7.15	0.1	5.4	0.72	0.69	3.7	0.2	3.1	0.42
ANFIS	Combination 1	0.9	0.73	9.21	0.12	6.96	0.54	0.68	6.09	0.32	4.77	0.01
	Combination 2	0.9	0.88	6.42	0.09	4.52	0.78	0.68	5.61	0.3	4.75	0.01
	Combination 3	0.9	0.9	5.81	0.08	4.04	0.82	0.58	4.3	0.23	3.43	0.25
	Combination 4	0.9	0.9	5.78	0.08	4.14	0.82	0.71	4.14	0.22	3.28	0.3
	Combination 5	0.9	0.89	6.06	0.08	4.47	0.8	0.63	5.25	0.28	4.14	0.1
	Combination 6	0.9	0.91	5.47	0.07	4	0.84	0.64	4.77	0.25	3.79	0.08
	Combination 7	0.9	0.91	5.52	0.07	4.1	0.83	0.5	4.53	0.24	3.9	0.17
GEP	Combination 1	-	0.71	9.54	0.13	7.23	0.5	0.66	8.01	0.37	6.79	0.01
	Combination 2	-	0.74	9.12	0.12	6.98	0.54	0.73	10.67	0.5	9.37	0.01
	Combination 3	-	0.78	9.8	0.13	7.61	0.47	0.74	8.05	0.37	7.16	0.01
	Combination 4	-	0.8	8.17	0.11	6.16	0.63	0.86	5.87	0.27	4.97	0.01
	Combination 5	-	0.55	11.3	0.15	8.58	0.3	0.55	21.49	1	20.51	0.01
	Combination 6	-	0.81	9.79	0.13	7.92	0.47	0.9	4.14	0.19	3.3	0.45
	Combination 7	-	0.79	8.23	0.11	6.23	0.63	0.62	7.93	0.37	6.49	0.01
GMDH	Combination 1	15,15,1	0.84	8.24	0.1	6.01	0.7	0.79	8.6	0.12	6.64	0.61
	Combination 2	15,15,1	0.91	6.57	0.08	4.72	0.83	0.89	5.1	0.11	3.7	0.76
	Combination 3	15,15,1	0.92	5.84	0.07	4.21	0.84	0.88	7.14	0.1	4.74	0.77
	Combination 4	15,15,1	0.92	5.77	0.07	3.99	0.85	0.9	6.18	0.11	4.21	0.82
	Combination 5	15,15,1	0.91	5.68	0.07	4.34	0.84	0.85	8.27	0.12	6.45	0.71
	Combination 6	15,15,1	0.93	5.28	0.07	3.63	0.87	0.93	4.37	0.07	3.87	0.83
	Combination 7	15,15,1	0.92	5.65	0.07	4.33	0.85	0.87	7.98	0.11	5.54	0.76
LSSVM	Combination 1	10,5	0.75	8.95	0.12	6.83	0.56	0.66	6.46	0.34	5.33	0.01
	Combination 2	10,5	0.91	5.72	0.08	4.15	0.82	0.73	6.33	0.33	5.55	0.01
	Combination 3	10,5	0.92	5.21	0.07	3.79	0.85	0.75	5.77	0.31	4.9	0.01
	Combination 4	10,5	0.92	5.43	0.07	3.96	0.84	0.79	5.54	0.29	4.7	0.01
	Combination 5	10,5	0.91	5.69	0.08	4.26	0.82	0.64	6.18	0.33	5.25	0.01
	Combination 6	10,5	0.92	5.05	0.07	3.65	0.86	0.8	5.74	0.3	4.88	0.01
	Combination 7	10,5	0.92	5.31	0.07	3.99	0.85	0.57	7.68	0.41	6.64	0.01

It should be remarked that combinations are categorized into three groups in the present application (Table 1), which (1) hydrogeological based: and employs only Qanat discharge data (QD_{t-1}, QD_{t-2}, QD_{t-3}) to predict QD (combination 1); (2) mixture of hydroclimatic and hydrogeological based: employs monthly local QD and QD of neighboring Qanat, local GWL, temperature, (T), precipitation (P), and evapotranspiration (ET) to predict QD (combinations 2, 3, 4, 6); And, (3) hydrological and hydrogeological based combinations without using previous QD values to predict QD (combination 5 and 7).

Models for QD prediction could be categorized into various groups based on NSE (Nash-Sutcliffe Efficiency) as follows: very good ($0.85 < \text{NSE} \le 1.00$), good ($0.70 < \text{NSE} \le 0.85$), satisfactory ($0.55 < \text{NSE} \le 0.70$), and unsatisfactory ($\text{NSE} \le 0.55$) (Moriasi et al. 2015). All models except the GEP model in the training phase for combinations 5 and 7 show satisfactory and good results according to NSE results ($0.55 < \text{NSE} \le 0.70$ and $0.70 < \text{NSE} \le 0.85$). Tables 2 and 5 show that the GMDH model gives the best results among other applied models for combinations 5 and 7 for the training and testing phases of both Qanats. For the one month ahead QD prediction in Qanat 1, the NSE values in the training phase are 0.84 (combination 5) and 0.85 (combination 7), and in Qanat 2, the corresponding values

are 0.82 and 0.83. Hence, GMDH can get good estimation without using previous QD values, and this model can be efficient for predicting QD when monthly QD time series data are unavailable.

Also, the conclusions reveal that considering QD of adjacent Qanat as input model parameters (comparing combination 4 with 6 and 5 with 7) improves the accuracy of all models. For example, in the training phase of LSSVM for the prediction of one month ahead QD in Qanat 1, from combination 4 to combination 6, the NSE is improved from 0.84 to 0.86 and from combination 5 to combination 7, the NSE is improved from 0.82 to 0.85; for the prediction of one month ahead QD in Qanat 2, from combination 4 combination 6, the NSE is improved from 0.76 to 0.80; and from combination 5 to combination 7, the NSE is improved from 0.75 to 0.79.

It is essential to point out that all models' results for combination 1, where just monthly QD (QD_{t-1}, QD_{t-2}, QD_{t-3}) values are used as inputs are not satisfactory for both Qanats. By considering all combinations and ML models like ANN, ANFIS, GEP, and LSSVM, it can be seen that combination six (combination with hydroclimatic and hydrogeological based) can offer accurate predictions for one month ahead QD. Based on Tab.s 2 and 5, it is evident that ANFIS, GMDH, and LSSVM models at the training phase offer good

Fig. 5 The observed and simulated one month ahead QD time series using the ANN, ANFIS, GEP, GMDH and LSSVM models for the combination 6: Qanat1 (**a**) and Qanat2 (**b**)
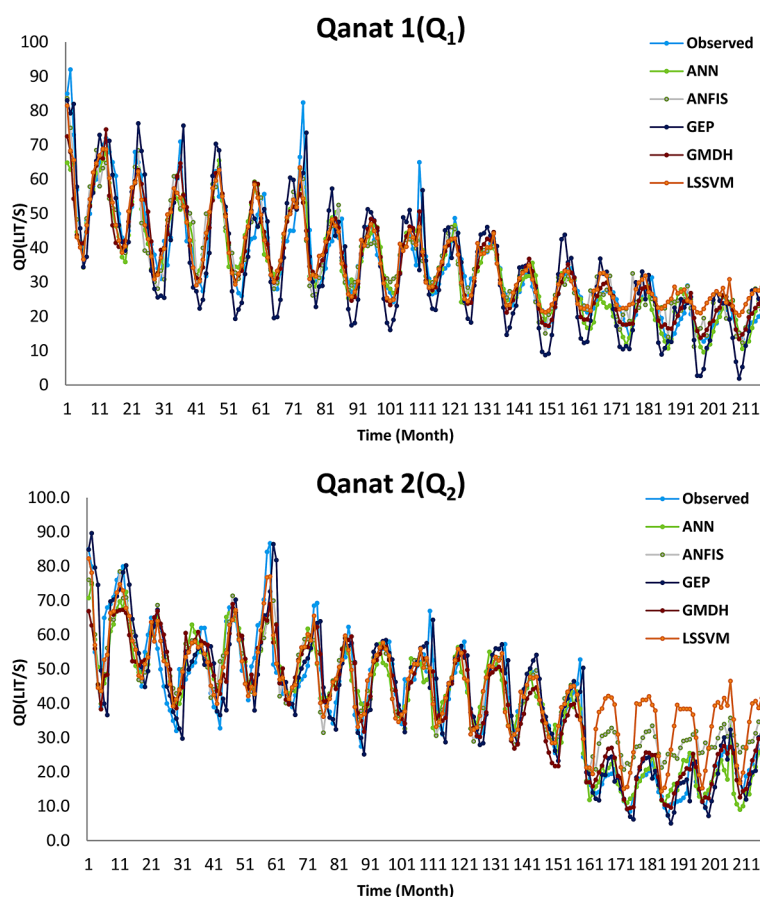
Table 3 The results of the different combinations in predicting two-months ahead QD for Qanat 1

Qanat 1	Input	Properties	Lt+2 Training					Test				
			R	RMSE	NRMSE	MAE	NS	R	RMSE	NRMSE	MAE	NS
ANN	Combination 1	2	0.52	11.12	0.17	8.9	0.26	0.58	4.07	0.22	3.33	0.31
	Combination 2	3	0.81	7.56	0.12	5.9	0.66	0.71	3.48	0.18	2.86	0.5
	Combination 3	4	0.82	7.44	0.11	5.69	0.67	0.73	3.88	0.21	3.06	0.37
	Combination 4	5	0.81	7.54	0.12	6.04	0.66	0.74	3.49	0.18	2.95	0.49
	Combination 5	6	0.79	7.9	0.12	6.07	0.63	0.29	4.93	0.26	3.87	0.1
	Combination 6	6	0.82	7.47	0.11	5.75	0.67	0.74	3.28	0.17	2.59	0.55
	Combination 7	0.5	0.8	7.29	0.12	6	0.65	0.7	3.69	0.2	2.9	0.43
ANFIS	Combination 1	0.9	0.53	11.03	0.17	8.58	0.28	0.42	9.36	0.5	7.98	0.01
	Combination 2	0.9	0.84	7.07	0.11	5.43	0.7	0.52	8.13	0.43	7.08	0.01
	Combination 3	0.9	0.86	6.55	0.1	4.93	0.75	0.63	6.31	0.33	5.32	0.01
	Combination 4	0.9	0.87	6.35	0.1	4.79	0.76	0.63	5.77	0.31	4.79	0.01
	Combination 5	0.9	0.88	6.08	0.09	4.55	0.78	0.65	5.42	0.29	4.09	0.01
	Combination 6	0.9	0.89	5.95	0.09	4.49	0.79	0.76	8.5	0.45	7.39	0.01
	Combination 7	0.9	0.89	6.07	0.09	4.48	0.79	0.7	5.48	0.29	4.43	0.01
GEP	Combination 1	-	0.46	11.48	0.18	9.27	0.21	0.23	11.69	0.54	10.25	0.01
	Combination 2	-	0.68	9.66	0.15	7.71	0.44	0.8	11.31	0.53	10.76	0.01
	Combination 3	-	0.67	9.64	0.15	7.81	0.44	0.76	15.51	0.72	14.51	0.01
	Combination 4	-	0.72	8.99	0.14	7.3	0.51	0.82	8.13	0.38	6.97	0.01
	Combination 5	-	0.61	10.27	0.16	8.11	0.37	0.72	19.72	0.92	18.76	0.01
	Combination 6	-	0.73	8.91	0.14	7.05	0.52	0.81	7.16	0.33	5.83	0.01
	Combination 7	-	0.7	9.27	0.14	7.02	0.48	0.59	6.19	0.29	5.01	0.01
GMDH	Combination 1	15,15,1	0.73	9.71	0.16	7.66	0.54	0.62	11.37	0.16	8.85	0.37
	Combination 2	15,15,1	0.9	6.03	0.1	4.64	0.8	0.85	8.38	0.12	5.95	0.71
	Combination 3	15,15,1	0.9	6.02	0.09	4.63	0.8	0.89	7.07	0.13	5.15	0.77
	Combination 4	15,15,1	0.9	6	0.09	4.55	0.8	0.89	6.72	0.12	5.22	0.8
	Combination 5	15,15,1	0.9	6.39	0.09	4.74	0.81	0.86	6.94	0.13	5.48	0.7
	Combination 6	15,15,1	0.91	5.9	0.08	4.1	0.81	0.89	6.49	0.11	4.86	0.8
	Combination 7	15,15,1	0.9	6.53	0.09	4.97	0.81	0.88	6.69	0.12	4.94	0.77
LSSVM	Combination 1	10,5	0.56	10.78	0.16	8.49	0.31	0.31	9.12	0.48	7.37	0.01
	Combination 2	10,5	0.86	6.71	0.1	5.07	0.73	0.65	7.11	0.38	6.17	0.01
	Combination 3	10,5	0.87	6.48	0.1	4.9	0.75	0.64	6.77	0.36	5.76	0.01
	Combination 4	10,5	0.89	6.04	0.09	4.52	0.78	0.64	6.88	0.36	5.89	0.01
	Combination 5	10,5	0.87	6.36	0.1	4.79	0.76	0.69	6.27	0.33	5.27	0.01
	Combination 6	10,5	0.91	5.4	0.08	3.88	0.83	0.71	7.15	0.38	6.29	0.01
	Combination 7	10,5	0.88	6.13	0.09	4.59	0.78	0.71	7.82	0.41	6.92	0.01

Fig. 6 Scatterplots of the observed and predicted one month ahead QD by different models for the combination 6: Qanat1 (left panel) and Qanat2 (Right panel)
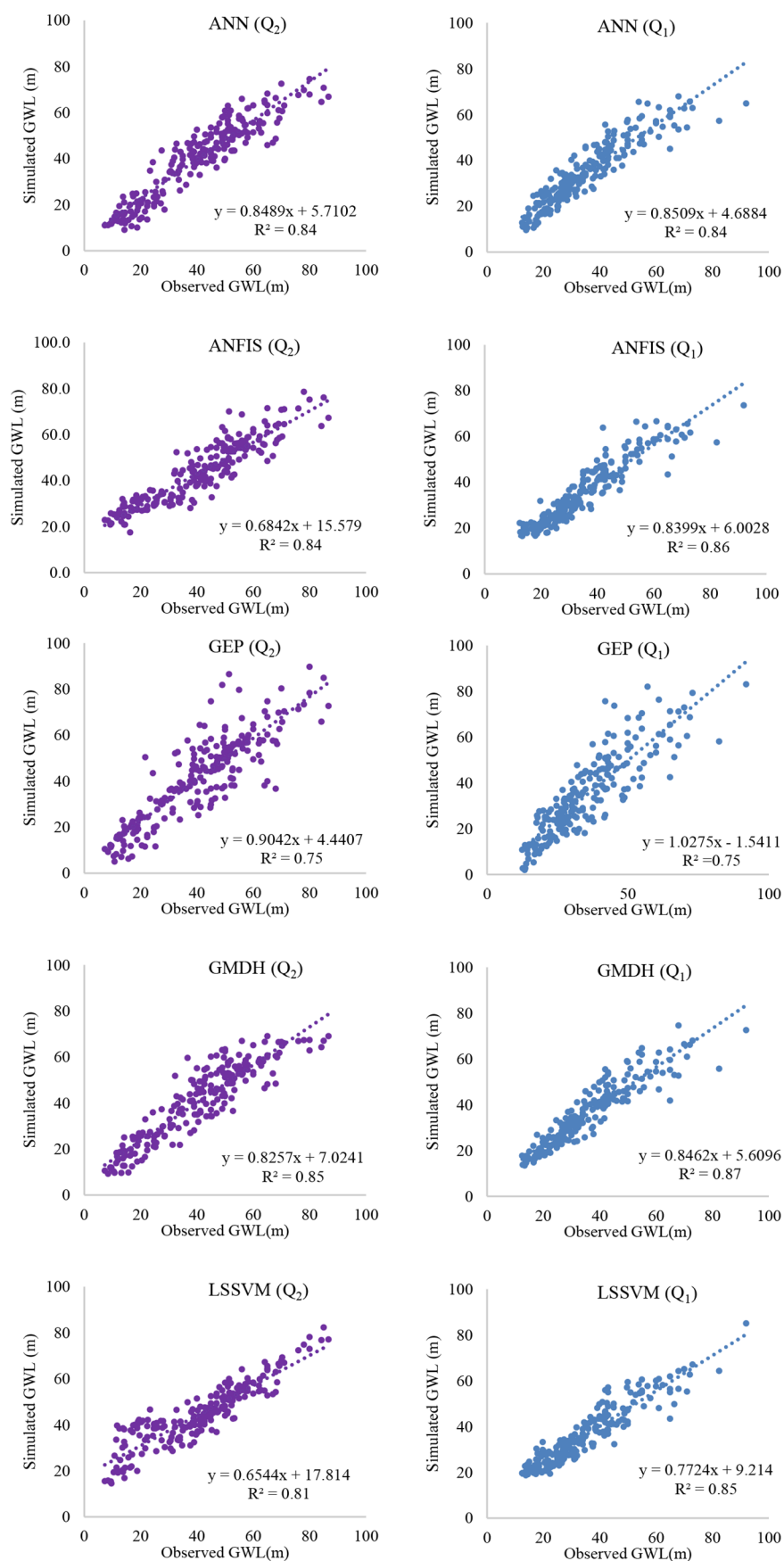
Table 4 The results of the different combinations in predicting five-months ahead QD for Qanat 1

Qanat 1	Input	Properties	Lt+5 Training					Test				
			R	RMSE	NRMSE	MAE	NS	R	RMSE	NRMSE	MAE	NS
ANN	Combination 1	2	0.35	11.96	0.18	9.62	0.11	0.16	4.93	0.26	4.06	0.01
	Combination 2	3	0.8	7.57	0.12	6.07	0.64	0.37	4.86	0.26	3.89	0.02
	Combination 3	4	0.81	7.36	0.11	5.94	0.66	0.51	4.87	0.26	3.88	0.01
	Combination 4	5	0.81	7.43	0.11	5.91	0.66	0.4	4.87	0.26	3.94	0.01
	Combination 5	6	0.8	7.48	0.11	5.48	0.65	0.15	7.65	0.4	5.82	0.01
	Combination 6	6	0.82	7.5	0.11	5.72	0.69	0.44	6.4	0.34	5.21	0.01
	Combination 7	0.5	0.81	7.39	0.11	5.52	0.66	0.14	5.73	0.3	4.84	0.01
ANFIS	Combination 1	0.9	0.37	11.86	0.18	9.31	0.14	0.12	10.7	0.57	8.77	0.01
	Combination 2	0.9	0.81	7.53	0.12	5.86	0.65	0.38	8.49	0.45	7.14	0.01
	Combination 3	0.9	0.81	7.48	0.11	5.79	0.66	0.38	8.09	0.43	6.78	0.01
	Combination 4	0.9	0.83	7.08	0.11	5.56	0.69	0.37	7.85	0.42	6.6	0.01
	Combination 5	0.9	0.85	6.65	0.11	5.33	0.73	0.66	5.7	0.3	4.67	0.01
	Combination 6	0.9	0.86	6.56	0.1	5.17	0.74	0.66	14.55	0.77	12.93	0.01
	Combination 7	0.9	0.86	6.58	0.1	5.23	0.73	0.67	5.97	0.32	4.84	0.01
GEP	Combination 1	-	0.29	12.31	0.19	9.93	0.06	0.36	19.52	0.91	18.83	0.01
	Combination 2	-	0.57	10.67	0.16	8.56	0.29	0.64	16.26	0.76	15.63	0.01
	Combination 3	-	0.39	11.73	0.18	9.5	0.14	0.52	14.06	0.65	13.1	0.01
	Combination 4	-	0.57	10.45	0.16	8.51	0.32	0.7	14.09	0.66	12.89	0.01
	Combination 5	-	0.64	9.93	0.15	7.83	0.39	0.65	8.19	0.38	7.25	0.01
	Combination 6	-	0.56	10.73	0.16	8.55	0.28	0.67	17.42	0.81	16.86	0.01
	Combination 7	-	0.53	10.85	0.17	8.55	0.27	0.65	18.39	0.86	17.8	0.01
GMDH	Combination 1	15,15,1	0.61	11.31	0.16	8.84	0.37	0.5	11.57	0.21	9.86	0.25
	Combination 2	15,15,1	0.81	8.23	0.12	6.58	0.65	0.8	8.56	0.15	6.91	0.64
	Combination 3	15,15,1	0.84	7.35	0.12	5.85	0.7	0.82	9.66	0.14	7.25	0.62
	Combination 4	15,15,1	0.85	7.5	0.11	5.79	0.71	0.81	8.25	0.15	6.5	0.66
	Combination 5	15,15,1	0.83	7.67	0.11	6.09	0.69	0.8	8.91	0.15	6.9	0.64
	Combination 6	15,15,1	0.86	7.27	0.1	5.36	0.73	0.89	6.44	0.11	5.14	0.79
	Combination 7	15,15,1	0.84	7.98	0.11	6.26	0.7	0.82	7.26	0.14	5.73	0.65
LSSVM	Combination 1	10,5	0.38	11.83	0.18	9.34	0.14	0.05	11.5	0.61	9.61	0.01
	Combination 2	10,5	0.81	7.49	0.11	5.89	0.66	0.34	8.3	0.44	7.22	0.01
	Combination 3	10,5	0.82	7.26	0.11	5.69	0.68	0.3	8.13	0.43	7.06	0.01
	Combination 4	10,5	0.85	6.84	0.1	5.4	0.71	0.34	8.35	0.44	7.26	0.01
	Combination 5	10,5	0.85	6.85	0.1	5.25	0.71	0.62	6.62	0.35	5.76	0.01
	Combination 6	10,5	0.87	6.39	0.1	5.01	0.75	0.64	9.21	0.49	8.41	0.01
	Combination 7	10,5	0.86	6.62	0.1	5.09	0.73	0.71	8.35	0.44	7.62	0.01

Table 5 The results of the different combinations in predicting one- ahead QD for Qanat 2

| Qanat 2 | Input | Properties | Lt+1 | | | | | | | | | |
| | | | Training | | | | | Test | | | | |
			R	RMSE	NRMSE	MAE	NS	R	RMSE	NRMSE	MAE	NS
ANN	Combination 1	2	0.66	9.42	0.15	7.14	0.43	0.6	4.67	0.22	3.66	0.3
	Combination 2	3	0.79	7.93	0.13	6.23	0.6	0.72	4.25	0.2	3.36	0.42
	Combination 3	4	0.8	7.81	0.13	5.85	0.61	0.75	4.26	0.2	3.36	0.42
	Combination 4	5	0.82	7.26	0.12	5.54	0.66	0.64	4.66	0.22	3.59	0.31
	Combination 5	6	0.81	7.3	0.12	5.69	0.66	0.45	5.33	0.25	4.38	0.09
	Combination 6	6	0.83	7.09	0.12	5.56	0.68	0.63	4.61	0.22	3.7	0.32
	Combination 7	0.5	0.82	7.43	0.12	5.8	0.65	0.48	5.03	0.24	4.2	0.19
ANFIS	Combination 1	0.9	0.67	9.9	0.14	7.38	0.45	0.51	8.17	0.39	6.58	0.01
	Combination 2	0.9	0.83	7.56	0.1	5.81	0.68	0.68	13.88	0.65	13.21	0.01
	Combination 3	0.9	0.85	7.03	0.1	5.36	0.72	0.68	13.95	0.66	11.6	0.01
	Combination 4	0.9	0.86	6.79	0.09	5.14	0.74	0.7	12.37	0.58	10.22	0.01
	Combination 5	0.9	0.86	6.87	0.09	5.17	0.74	0.69	15.87	0.75	14.85	0.01
	Combination 6	0.9	0.88	6.01	0.08	4.63	0.76	0.74	11.26	0.53	10.6	0.01
	Combination 7	0.9	0.87	6.7	0.09	5.07	0.75	0.67	18.78	0.89	17.44	0.01
GEP	Combination 1	-	0.64	9.67	0.16	7.26	0.4	0.75	8.39	0.18	5.97	0.18
	Combination 2	-	0.68	9.36	0.15	7.57	0.44	0.63	26.86	0.59	0.01	0.01
	Combination 3	-	0.63	10.82	0.18	8.23	0.25	0.77	6.62	0.15	4.32	0.49
	Combination 4	-	0.68	10.44	0.17	7.61	0.3	0.63	13.27	0.29	11.68	0.01
	Combination 5	-	0.61	9.94	0.16	7.83	0.37	0.54	30.41	0.67	29.18	0.01
	Combination 6	-	0.69	10.29	0.17	7.93	0.39	0.78	6.66	0.15	4.8	0.48
	Combination 7	-	0.68	9.82	0.16	7.6	0.38	0.6	14.06	0.31	12.43	0.01
GMDH	Combination 1	15,15,1	0.86	9	0.11	6.65	0.73	0.85	9.96	0.13	7.48	0.71
	Combination 2	15,15,1	0.91	7.53	0.09	5.66	0.83	0.86	8.19	0.11	5.94	0.72
	Combination 3	15,15,1	0.91	7.32	0.09	6.05	0.83	0.88	8.5	0.11	6.64	0.77
	Combination 4	15,15,1	0.92	7.18	0.09	5.48	0.84	0.88	8.91	0.11	7.11	0.78
	Combination 5	15,15,1	0.91	7.69	0.1	5.72	0.82	0.88	7.88	0.1	5.64	0.78
	Combination 6	15,15,1	0.93	6.56	0.09	5.13	0.86	0.89	8.06	0.12	6.01	0.79
	Combination 7	15,15,1	0.91	7.06	0.1	5.46	0.83	0.89	9.27	0.12	7.11	0.79
LSSVM	Combination 1	10,5	0.68	9.8	0.13	7.39	0.46	0.55	8.33	0.39	6.88	0.01
	Combination 2	10,5	0.85	7.01	0.1	5.35	0.73	0.78	14.08	0.66	13.42	0.01
	Combination 3	10,5	0.86	6.81	0.09	5.19	0.74	0.78	12.37	0.58	11.67	0.01
	Combination 4	10,5	0.87	6.58	0.09	5.02	0.76	0.75	13.36	0.63	12.59	0.01
	Combination 5	10,5	0.86	6.63	0.09	5.1	0.75	0.66	16.04	0.76	14.67	0.01
	Combination 6	10,5	0.89	6.05	0.08	4.59	0.8	0.73	14.59	0.69	13.74	0.01
	Combination 7	10,5	0.88	6.15	0.08	4.61	0.79	0.67	16	0.75	14.84	0.01

precision for the one month ahead QD prediction for both Qanats with NSE higher than 0.55. However, great implementation is found for the GMDH methods based on NSE values greater than 0.7 for five months ahead in the training phase for both Qanats. Considering RMSE, NRMSE and R values, the GMDH model reveals precise results with the low RMSE, NRMSE, and high R values. The LSSVM can be ranked as the second model. According to the NSE analysis, the GEP model is the worst in estimating QD. The outcomes of the training phase ascertained that the GMDH model (six input combination) for both Qanats performed superior to the other models; for one month ahead QD prediction in Qanat 1, $R=0.93$, RMSE $=5.28$, NRMSE $=0.07$, MAE $=3.63$ and NSE $=0.87$; for two-month ahead QD prediction in Qanat 1, $R=0.91$, RMSE $=5.90$, NRMSE $=0.08$,

MAE $=4.10$ and NSE $=0.81$; for five-month ahead QD prediction in Qanat 1, $R=0.86$, RMSE $=7.27$, NRMSE $=0.10$, MAE $=5.36$ and NSE $=0.73$; for one month ahead QD prediction in Qanat 2, $R=0.93$, RMSE $=6.56$, NRMSE $=0.09$, MAE $=5.13$ and NSE $=0.86$; for two-month ahead QD prediction in Qanat 2, $R=0.89$, RMSE $=7.71$,NRMSE $=0.10$, MAE $=6.28$ and NSE $=0.79$; for five-month ahead QD prediction in Qanat 2, $R=0.86$, RMSE $=9.11$,NRMSE $=0.11$, MAE $=7.51$ and NSE $=0.73$)

The accuracy models for the combination six are further compared in Figs. 5 and 6 via time variation graphs and scatterplots for both Qanats. The graphs in Fig. 5 provide detailed changes between observed values and predictions provided by the models. Also, the provided graphs in Fig. 6 reveal how all model's simulations are scattering,

Table 6 The results of the different combinations in predicting two-months ahead QD for Qanat 2

Qanat 2	Input	Properties	Lt+2									
			Training					Test				
			R	RMSE	NRMSE	MAE	NS	R	RMSE	NRMSE	MAE	NS
ANN	Combination 1	2	0.44	11.05	0.18	8.8	0.19	0.38	5.43	0.25	4.6	0.12
	Combination 2	3	0.72	8.69	0.14	6.59	0.5	0.44	5.74	0.26	4.76	0.01
	Combination 3	4	0.75	8.23	0.13	6.27	0.55	0.45	5.43	0.25	4.45	0.11
	Combination 4	5	0.75	8.33	0.14	6.44	0.54	0.54	5.21	0.24	4.32	0.19
	Combination 5	6	0.73	9.59	0.16	7.73	0.39	0.43	5.85	0.27	4.83	0.01
	Combination 6	6	0.75	8.37	0.14	6.58	0.53	0.62	5.04	0.23	4.15	0.24
	Combination 7	0.5	0.74	8.68	0.14	6.9	0.5	0.48	6.13	0.28	4.8	0.01
ANFIS	Combination 1	0.9	0.47	11.77	0.16	8.94	0.22	0.22	9.14	0.42	6.98	0.01
	Combination 2	0.9	0.78	8.42	0.12	6.63	0.6	0.64	17.45	0.8	16.43	0.01
	Combination 3	0.9	0.81	7.92	0.11	6.28	0.65	0.62	19.98	0.92	18.11	0.01
	Combination 4	0.9	0.82	7.64	0.1	6.1	0.67	0.63	16.35	0.75	13.93	0.01
	Combination 5	0.9	0.83	7.33	0.1	5.74	0.7	0.68	17.22	0.79	14.36	0.01
	Combination 6	0.9	0.84	7.24	0.1	5.6	0.71	0.67	14.66	0.68	11.88	0.01
	Combination 7	0.9	0.84	7.3	0.1	5.66	0.7	0.69	16.64	0.77	13.94	0.01
GEP	Combination 1	-	0.41	11.21	0.18	8.78	0.16	0.54	22.51	0.49	21.24	0.01
	Combination 2	-	0.6	10.31	0.17	8.17	0.32	0.58	10.81	0.24	8.13	0.01
	Combination 3	-	0.59	10	0.16	7.64	0.33	0.67	24.97	0.55	24.01	0.01
	Combination 4	-	0.58	10.02	0.16	7.75	0.33	0.61	22.6	0.5	19.69	0.01
	Combination 5	-	0.58	10.02	0.16	7.73	0.33	0.53	29.26	0.64	28.17	0.01
	Combination 6	-	0.61	9.89	0.16	7.65	0.35	0.54	9.35	0.62	7.24	0.01
	Combination 7	-	0.6	9.83	0.16	7.56	0.36	0.62	24.42	0.54	23.31	0.01
GMDH	Combination 1	15,15,1	0.81	10.79	0.14	8.04	0.65	0.71	12.2	0.18	10.04	0.4
	Combination 2	15,15,1	0.87	8.53	0.12	7.07	0.76	0.82	10.04	0.14	8.33	0.67
	Combination 3	15,15,1	0.87	8.9	0.11	7.29	0.75	0.84	9.49	0.15	7.32	0.68
	Combination 4	15,15,1	0.88	8.52	0.11	6.9	0.77	0.86	9.34	0.14	8.2	0.71
	Combination 5	15,15,1	0.89	7.77	0.1	6.43	0.81	0.82	10.09	0.13	8.19	0.65
	Combination 6	15,15,1	0.89	7.71	0.1	6.28	0.79	0.88	8.64	0.13	8.12	0.72
	Combination 7	15,15,1	0.89	8.38	0.11	6.79	0.79	0.84	8.54	0.12	6.72	0.69
LSSVM	Combination 1	10,5	0.52	11.46	0.16	8.75	0.26	0.26	10.07	0.46	7.72	0.01
	Combination 2	10,5	0.8	8	0.11	6.34	0.64	0.67	18.64	0.86	17.52	0.01
	Combination 3	10,5	0.81	7.85	0.11	6.17	0.65	0.65	15.59	0.72	14.49	0.01
	Combination 4	10,5	0.83	7.57	0.1	5.96	0.68	0.61	15.33	0.71	14.38	0.01
	Combination 5	10,5	0.83	7.48	0.1	5.87	0.69	0.65	16.12	0.74	14.66	0.01
	Combination 6	10,5	0.84	7.21	0.1	5.77	0.71	0.6	16.08	0.74	15.18	0.01
	Combination 7	10,5	0.84	7.23	0.1	5.79	0.71	0.67	16.54	0.76	15.29	0.01

and R values provide a practical understanding of fitting the simulated dataset based on the accuracy and precision of both Qanats. Analysis of the hydrographs and scatterplots reveals that the forecasted time series by GMDH closely aligns with the observed QD values and exhibits less dispersion compared to the other four models used. However, noticeable disparities between the predicted and observed time series are apparent for the GEP model. However, the GMDH model could not catch the extreme QD values. This can be explained by the smaller number of samples for QD peak values. Without a high number of examples, ML models cannot adequately learn extreme events.

The Taylor diagrams for both Qanat 1 and Qanat 2, as shown in Fig. 7, provide a visual comparison of various machine learning models based on their statistical performance metrics, including standard deviation, correlation, and root mean square error. In each diagram, the GMDH model is closer to the reference point, indicating that its predictions closely match the observed data's variability (standard deviation) and pattern (correlation). The proximity of the GMDH model to the center of the circle further demonstrates its lower RMSE, signifying more accurate predictions compared to other models. These results underscore the effectiveness of the GMDH model in capturing the complex dynamics of Qanat discharge, providing reliable predictions that are crucial for water resource management in the regions dependent on Qanats. The consistency in model performance across both Qanat 1 and Qanat 2 further reinforces the robustness of the GMDH approach.

Table 7 The results of the different combinations in predicting five-months ahead QD for Qanat 2

| Qanat 2 | Input | Properties | Lt+5 | | | | | | | | | |
| | | | Training | | | | | Test | | | | |
			R	RMSE	NRMSE	MAE	NS	R	RMSE	NRMSE	MAE	NS
ANN	Combination 1	2	0.44	10.83	0.18	8.21	0.19	0.11	7.33	0.32	5.67	0.01
	Combination 2	3	0.56	10.31	0.17	8.15	0.27	0.26	6.74	0.3	5.83	0.18
	Combination 3	4	0.59	9.76	0.16	7.79	0.34	0.29	7.13	0.31	6.1	0.23
	Combination 4	5	0.67	8.86	0.15	6.55	0.46	0.05	8.56	0.38	6.83	0.01
	Combination 5	6	0.65	9.3	0.15	7.54	0.4	0.2	6.35	0.28	4.96	0.01
	Combination 6	6	0.68	8.82	0.14	6.92	0.46	0.43	5.74	0.25	4.95	0.07
	Combination 7	0.5	0.68	9.02	0.15	7.18	0.44	0.28	7.29	0.32	5.75	0.07
ANFIS	Combination 1	0.9	0.38	12.35	0.17	9.41	0.14	0.05	11.07	0.49	8.63	0.01
	Combination 2	0.9	0.7	9.5	0.13	7.64	0.49	0.52	21.2	0.93	19.81	0.01
	Combination 3	0.9	0.79	8.16	0.11	6.43	0.63	0.47	20.46	0.9	16.75	0.01
	Combination 4	0.9	0.82	7.66	0.11	5.97	0.67	0.51	14.35	0.63	12.85	0.01
	Combination 5	0.9	0.82	7.66	0.11	5.94	0.67	0.47	20.46	0.9	16.75	0.01
	Combination 6	0.9	0.84	7.39	0.1	5.74	0.71	0.51	14.05	0.63	12.73	0.01
	Combination 7	0.9	0.83	7.34	0.1	5.64	0.7	0.46	19.46	0.9	16.15	0.01
GEP	Combination 1	-	0.24	11.69	0.19	9.09	0.06	0.04	45.23	0.99	37.45	0.01
	Combination 2	-	0.4	11.42	0.19	8.9	0.1	0.56	28.11	0.62	26.96	0.01
	Combination 3	-	0.41	10.97	0.18	8.56	0.17	0.49	26.22	0.58	24.99	0.01
	Combination 4	-	0.46	10.81	0.18	8.34	0.19	0.48	27.2	0.6	26.04	0.01
	Combination 5	-	0.39	11.08	0.18	9.2	0.15	0.1	21.98	0.48	20.65	0.01
	Combination 6	-	0.47	10.73	0.18	8.38	0.21	0.51	22.3	0.49	19.86	0.01
	Combination 7	-	0.45	10.81	0.18	8.34	0.19	0.48	27.2	0.6	26.04	0.01
GMDH	Combination 1	15,15,1	0.73	12.22	0.15	9.75	0.53	0.65	12.55	0.19	10.43	0.41
	Combination 2	15,15,1	0.77	10.78	0.15	8.89	0.6	0.72	13.18	0.17	10.84	0.48
	Combination 3	15,15,1	0.82	10.28	0.13	8.44	0.67	0.72	11.33	0.19	9.96	0.5
	Combination 4	15,15,1	0.82	9.89	0.13	8.1	0.68	0.84	9.71	0.12	7.93	0.7
	Combination 5	15,15,1	0.83	9.49	0.12	7.81	0.69	0.75	12.88	0.17	10.73	0.54
	Combination 6	15,15,1	0.86	9.11	0.11	7.51	0.73	0.71	12.47	0.21	10.36	0.45
	Combination 7	15,15,1	0.85	9.21	0.12	7.64	0.72	0.76	12.08	0.2	9.01	0.51
LSSVM	Combination 1	10,5	0.41	12.25	0.17	9.33	0.16	0.09	12.04	0.53	9.36	0.01
	Combination 2	10,5	0.74	9.07	0.12	7.24	0.54	0.45	21.49	0.95	19.71	0.01
	Combination 3	10,5	0.76	8.76	0.12	6.94	0.57	0.39	18.53	0.82	16.68	0.01
	Combination 4	10,5	0.78	8.44	0.12	6.69	0.6	0.35	17.59	0.78	16.05	0.01
	Combination 5	10,5	0.79	8.36	0.11	6.72	0.61	0.5	15.29	0.67	13.56	0.01
	Combination 6	10,5	0.8	8.17	0.11	6.49	0.62	0.36	18.15	0.8	16.73	0.01
	Combination 7	10,5	0.81	7.99	0.11	6.34	0.64	0.56	15.73	0.69	14.24	0.01

In total, the current study implies the superiority of the GMDH method compared to the other applied ML methods based on the employed dataset. These findings reinforce the prior results reported by the existing literature for different areas (Aghelpour and Varshavian 2020; Li et al. 2020; Moghadam et al. 2021; Kamali et al. 2022).

Figures 8 and 9 illustrate the impact of various hydrological, meteorological, and neighboring Qanat features on the discharge predictions of Qanats 1 and 2 using the GMDH model as the best model.

Figure 8 provides a global view, showcasing how different features influence the model output across all instances. This global perspective is crucial for understanding which features generally drive the model's predictions and should therefore be monitored closely in management practices.

QD1 and QD2 (discharging of neighboring Qanat) and Groundwater Level (GWL) emerge as significant drivers, with high variability in their SHAP values, indicating that these features have a substantial impact on the model's output across all data points. Temperature (T), Evapotranspiration (ET) and Precipitation (P) show less variability and smaller range of SHAP values, suggesting a lower but consistent influence.

Figure 9 delves into local interactions, revealing how specific feature combinations affect discharge outcomes under particular conditions. This analysis is crucial for recognizing circumstances where interactions between features like temperature and groundwater level critically alter predictions, offering deeper insights into the model's operation under diverse scenarios. According to this figure for both

Fig. 7 Taylor diagrams of different models: Qanat1 (**a**) and Qanat2 (**b**)
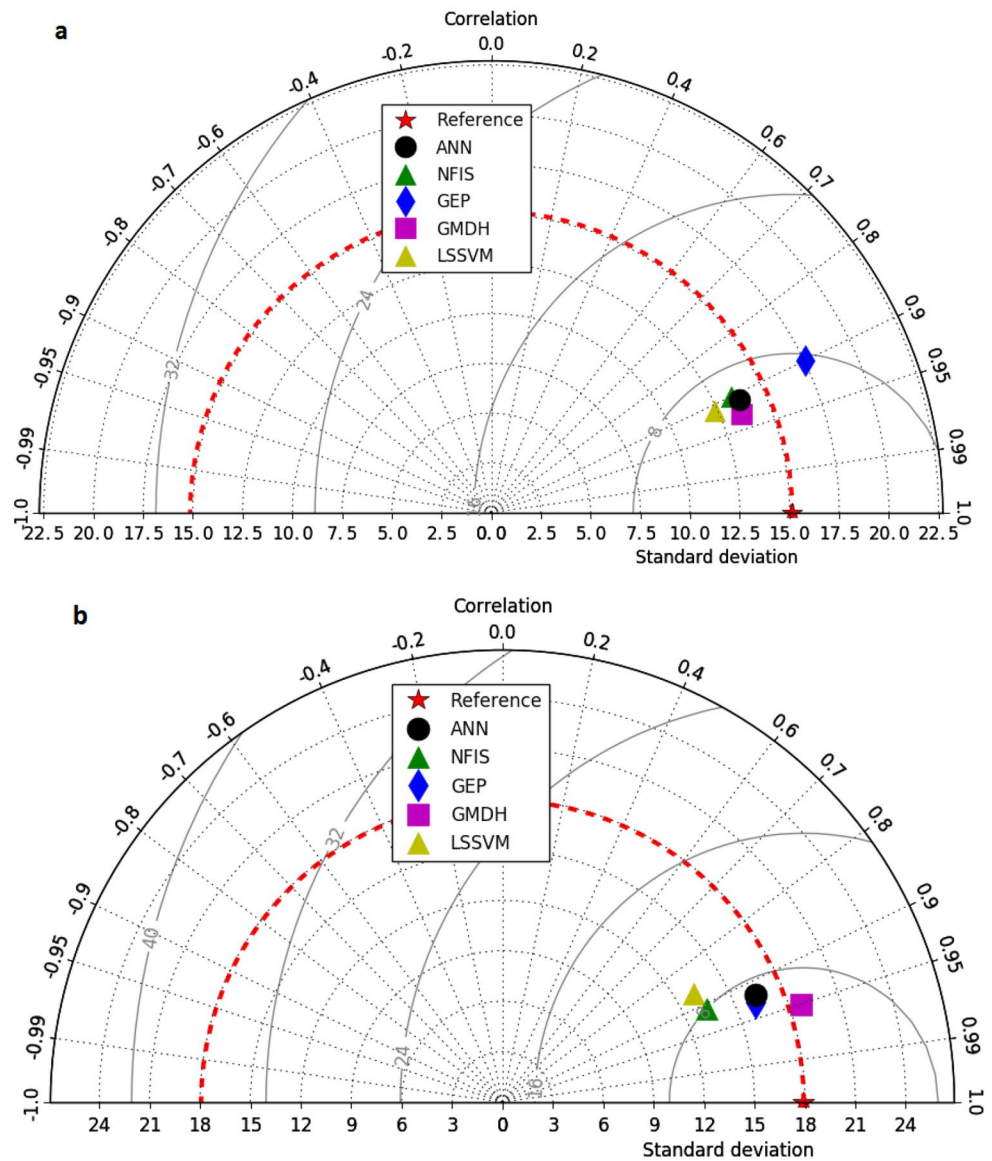
Qanats, the importance of temperature and significant local interaction with precipitation and GWL may indicate scenarios where climatic conditions jointly affect QD. Also, the interaction values for discharge from neighboring Qantas underscore the model's sensitivity to neighboring discharge influences, critical for precise discharge predictions.

The practical implications of the presented method for groundwater resource management in arid regions

The findings from this study, demonstrating the efficacy of various machine learning models in predicting Qanat discharge, have significant practical implications for groundwater resource management, particularly in arid regions where water scarcity is a pressing issue. Implementing these

models can enable local water authorities and stakeholders to forecast water availability more accurately, facilitating proactive management strategies. For instance, the predictive capabilities of these models can be integrated into water management systems to optimize the allocation and use of water from Qanats, ensuring that water supply meets the demands of agriculture and domestic use without overexploiting the aquifer. By predicting periods of low discharge, water managers can plan alternative water supply strategies or implement water-saving measures in advance, thus avoiding crises. Moreover, the ability of these models to account for various hydrological and climatic inputs means they can also be used to assess the impact of climate change on Qanat performance. This is crucial for long-term water resource planning, allowing for the adaptation of water infrastructure and policies in response to predicted changes
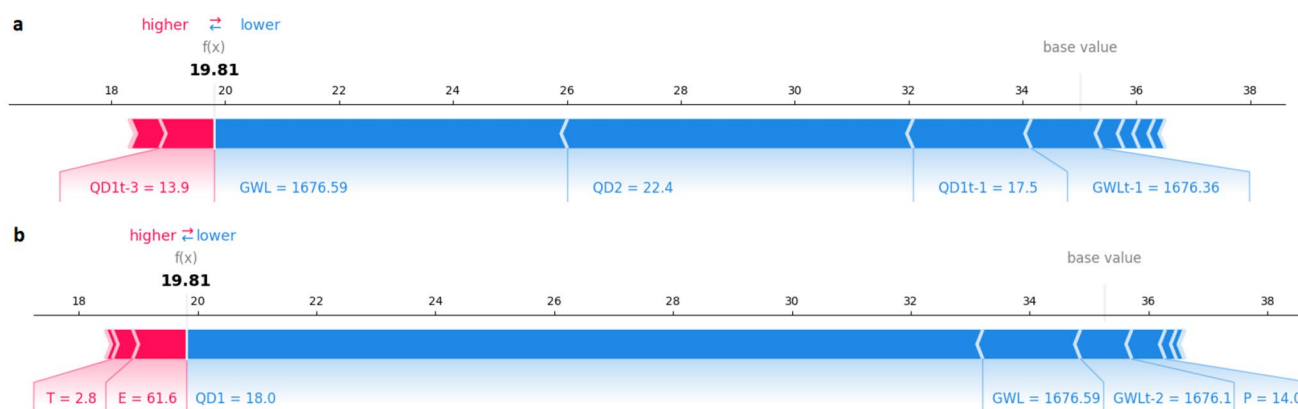
Fig. 8 Feature impact analysis on qanat discharge predictions for GMDH model: Qanat1 (**a**) and Qanat2 (**b**)

in groundwater levels and Qanat discharge patterns. Incorporating machine learning models into groundwater monitoring systems could also enhance the efficiency of these systems, providing continuous, real-time data analysis. This would allow for immediate responses to detected changes in water levels, potentially preventing over-extraction and its associated negative impacts on the aquifer.

Overall, the application of these models in groundwater resource management not only promises to enhance the sustainability of water resources in arid regions but also supports the preservation of Qanats, which are vital to the cultural and ecological landscapes of these areas. By fostering a more data-driven approach to water management, stakeholders can ensure the resilience and efficiency of water use in face of increasing variability due to human and environmental changes.

Also, these machine learning predictions help in mitigating adverse environmental effects. The environmental impacts of Qanat decline or alterations in their discharge patterns are profound, especially in arid and semi-arid regions where they serve as critical water sources. One significant impact is on local ecosystems, which rely on consistent and adequate water supply from Qanats. A decrease in water availability can lead to habitat degradation, loss of biodiversity, and reduced agricultural productivity due to increased soil salinity and desertification. This alteration in land use and ecosystem health can, in turn, reduce the land's natural resilience to environmental changes.

Machine learning models, as detailed in the paper, can significantly mitigate these adverse effects through enhanced predictive capabilities. By accurately forecasting Qanat discharge under various scenarios, these models allow for better water resource management. They enable authorities to implement strategic water allocation and conservation practices, ensuring that water usage is sustainable and does

not exceed the natural replenishment rates of aquifers. Additionally, these predictions can facilitate the development of early warning systems for water scarcity, allowing communities and farmers to implement water-saving measures in advance, thereby safeguarding against the ecological and economic impacts of drought.

Future research directions for enhancing Qanat discharge predictions using machine learning

For enhancing the accuracy of ML predictions in Qanat discharge studies, future research can explore several innovative approaches. Developing hybrid models that combine multiple ML techniques could leverage the strengths of various algorithms to handle complex datasets more effectively. Additionally, advanced ensemble techniques like stacked generalization could further refine predictions by integrating outputs from multiple models, exploiting the unique strengths of each to improve overall accuracy. Enhancing feature engineering and selection is also crucial, as more precise input features related to hydrological, meteorological, and geological data could be pivotal in improving model outcomes. Exploring deep learning architectures such as Convolutional Neural Networks (CNNs) for spatial data and Long Short-Term Memory (LSTM) networks for temporal data could offer significant advancements in processing complex time series and spatial patterns. Incorporating uncertainty analysis through Bayesian networks or Monte Carlo methods could provide deeper insights into prediction confidence and the influence of various parameters, which is vital for robust water management strategies. Additionally, expanding data sources to include remote sensing data could provide more dynamic and high-frequency updates, enhancing the model's responsiveness to environmental changes.
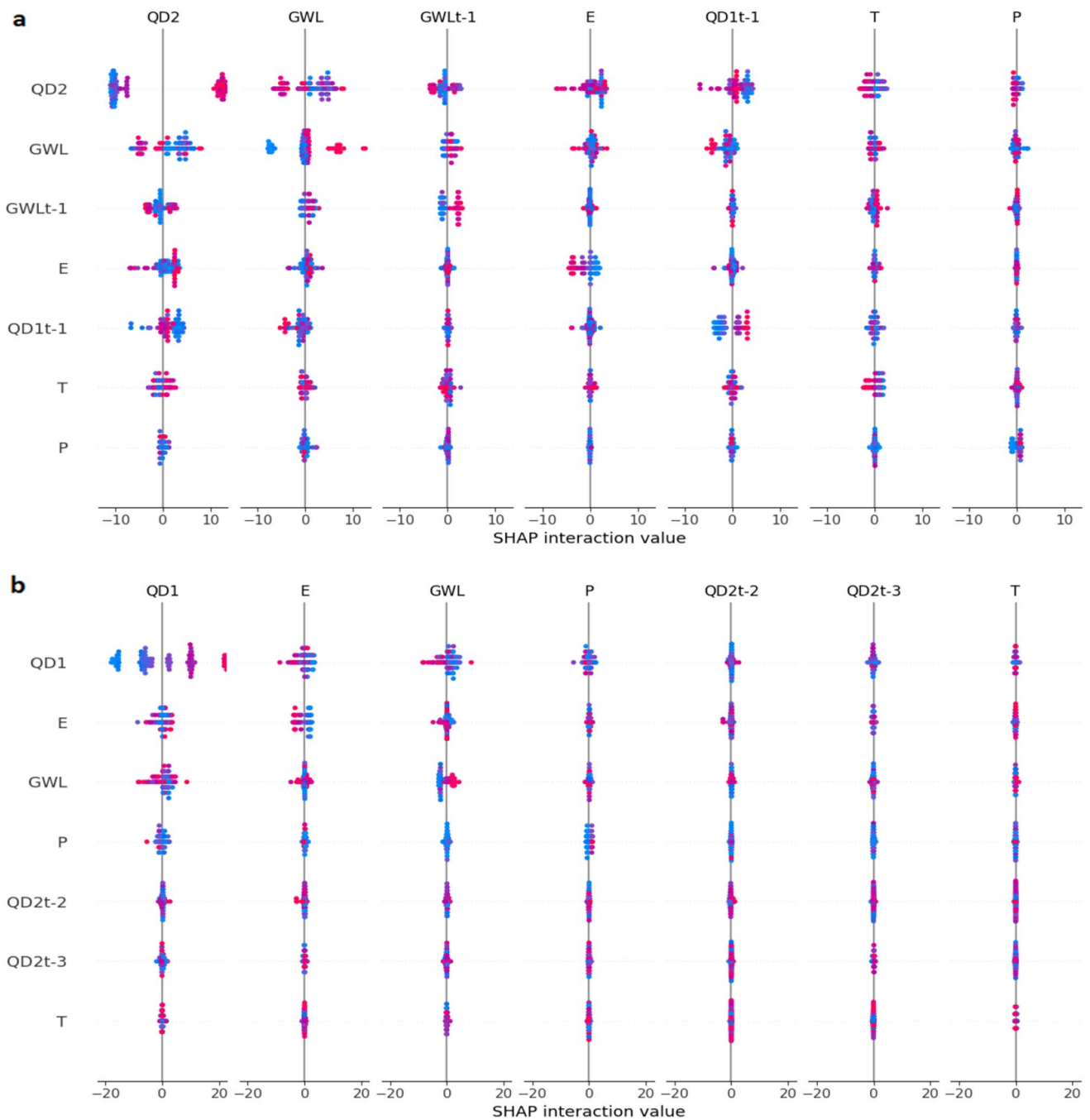
Fig. 9 SHAP interaction value analysis for Qanat discharge predictions using GMDH mode: Qanat1 (**a**) and Qanat2 (**b**)

Conclusions

Qanats are a unique way to extract groundwater from the aquifer, especially in arid and semiarid countries. Simulation of discharge of Qanats is essential for aquifer and demand management, and machine learning methods are promising tools to simulate the complex behavior of the Qanats. The present study utilized a range of well-accepted machine learning methods for predicting QD. Monthly local

QD (QD_{t-1}, QD_{t-2}, QD_{t-3}), QD of neighboring Qanat, local GWL for one, two, and three lag-times (GWL_{t-1}, GWL_{t-2}, GWL_{t-3}), temperature, (T), precipitation (P), and evapotranspiration (ET) data were considered as inputs to predict one-, two-, and five-month ahead QD for two adjacent Qanats in Razan-Ghahavand Aquifer. Seven input combinations were categorized into three groups based on expert knowledge and investigated for QD prediction with different lead times (one to five months ahead) using monthly

QD data from Razan- Ghahavand Aquifer. Combination 1 in group 1, with hydrogeology-based inputs, includes only Qanat discharge. Combinations 2, 3, 4 and 6 in group 2, with a mixture of hydroclimatic and hydrogeological based inputs, employ monthly QD, QD of neighboring Qanat, GWL, temperature, (T), precipitation (P), and evapotranspiration (ET) to predict QD; and combinations 5 and 7 in group 3, with hydroclimatic and hydrogeological based inputs, do not consider previous QD values. The results of the different models were investigated through different statistical error indices, R, RMSE, NRMSE, MAE, and NSE, to identify the superior model which could predict the QD and give satisfactory results. According to the NSE, the GMDH model gives the best results among the applied models for combinations 5 and 7 in-group 3 for both Qanats. Therefore, this model can efficiently predict QD when monthly QD time series data are unavailable. It is worth mentioning that the results of all used models for combination one, QDt-1, QDt-2, and QDt-3, were not satisfactory for both Qanats. Furthermore, the overall results indicate that considering QD of adjacent Qanat as an input for the developed models (comparing combination 4 with 6 and 5 with 7) generally increases the quality of the model's prediction. In general, based on the model results, the GMDH model had the superior results, followed by the LSSVM, ANFIS, and ANN, respectively, to predict QD in the present study, and the GEP model was not satisfactory for the simulation of QD.

Data availability Data will be available upon request.

Declarations

Competing interests The authors declare no competing interests.

Ethical approval Not applicable because this article does not contain any studies of human or animal subjects.

Consent to participate Not applicable.

Consent to publish The undersigned gives us consent to publish identifiable details in the journal, including text, figures, and tables.

Competing interests The authors declare that they have no competing interests.

References

Aghelpour P, Varshavian V (2020) Evaluation of stochastic and artificial intelligence models in modeling and predicting of river daily flow time series. Stoch Env Res Risk Assess 34(1):33–50. https://doi.org/10.1007/s00477-019-01761-4

Ahmadi A, Olyaei M, Heydari Z, Emami M, Zeynolabedin A, Ghomlaghi A, Sadegh M (2022) Groundwater Level modeling with machine learning: a systematic review and Meta-analysis. Water 14(6):949. https://doi.org/10.3390/w14060949

Antonopoulos VZ, Gianniou SK (2022) Analysis and modelling of temperature at the water–atmosphere interface of a Lake by Energy Budget and ANNs models. Environ Processes 9(1):1–20. https://doi.org/10.21203/rs.3.rs-843456/v1

Arya Azar N, Kayhomayoon Z, Ghordoyee Milan S, Zarif Sanayei H, Berndtsson R, Nematollahi Z (2022) A hybrid approach based on simulation, optimization, and estimation of conjunctive use of surface water and groundwater resources. Environ Sci Pollut Res 1–17. https://doi.org/10.1007/s11356-022-19762-2

Azimi H, Shiri H (2020) Ice-seabed interaction analysis in sand using a gene expression programming-based approach. Appl Ocean Res 98:102120. https://doi.org/10.1016/j.apor.2020.102120

Azari Rad M, Ziaei AN, Naghedifar MR (2018) Three-dimensional numerical modeling of submerged zone of Qanat hydraulics in unsteady conditions. J Hydrol Eng 23(3):04017063.

Bahmani R, Ouarda TB (2021) Groundwater level modeling with hybrid artificial intelligence techniques. J Hydrol 595:125659. https://doi.org/10.1016/j.jhydrol.2020.125659

Banadkooki FB, Ehteram M, Ahmed AN, Teo FY, Fai CM, Afan HA, El-Shafie A (2020) Enhancement of groundwater-level prediction using an integrated machine learning model optimized by whale algorithm. Nat Resour Res 29(5):3233–3252. https://doi.org/10.1007/s11053-020-09634-2

Boustani F (2008) Sustainable water utilization in arid region of Iran by Qanats. In Proceeding of world Academy of science, engineering and technology, 33, 213–216

Chang CC, Lin CJ (2011) LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol (TIST) 2(3):1–27. https://doi.org/10.1145/1961189.1961199

Cui F, Al-Sudani ZA, Hassan GS, Afan HA, Ahammed SJ, Yaseen ZM (2022) Boosted artificial intelligence model using improved alpha-guided grey wolf optimizer for groundwater level prediction: comparative study and insight for federated learning technology. J Hydrol 606:127384. https://doi.org/10.1016/j.jhydrol.2021.127384

Dehghani R, Poudeh T, H (2022) Application of novel hybrid artificial intelligence algorithms to groundwater simulation. Int J Environ Sci Technol 19(5):4351–4368. https://doi.org/10.1007/s13762-021-03596-5

Ebrahimi M, Sarikhani MR, Shiri J, Shahbazi F (2021) Modeling soil enzyme activity using easily measured variables: Heuristic alternatives. Appl Soil Ecol 157:103753.

Fels AEA, Ghorfi E, M (2022) Using remote sensing data for geological mapping in semi-arid environment: a machine learning approach. Earth Sci Inf 15(1):485–496

Ferreira C (2001) Gene expression programming: a new adaptive algorithm for solving problems. arXiv preprint cs/0102027. https://doi.org/10.48550/arXiv.cs/0102027

Ganjeizadeh Rohani F, Mohamadi N, Ganjei-Zadeh K (2024) Heavy metal distribution and assessment in Qanat system water sourced from the mountains surrounding the copper mine. Sustainable Water Resour Manage 10(3):107

Ghazi B, Jeihouni E, Kalantari Z (2021a) Predicting groundwater level fluctuations under climate change scenarios for Tasuj

plain, Iran. Arab J Geosci 14(2):1–12. https://doi.org/10.1007/s12517-021-06508-6

Ghazi B, Jeihouni E, Kouzehgar K, Haghighi AT (2021b) Assessment of probable groundwater changes under representative concentration pathway (RCP) scenarios through the wavelet–GEP model. Environ Earth Sci 80(12):1–15. https://doi.org/10.1007/s12665-021-09746-9

Goldberg DE (1989) Genetic algorithms in search, optimization, and machine learning. Addison Read. https://doi.org/10.5555/534133

Gu Y, Zhao W, Wu Z (2010) Least squares support vector machine algorithm [J]. J Tsinghua Univ (Science Technology) 7:1063–1066

Guzman SM, Paz JO, Tagert MLM, Mercer AE (2019) Evaluation of seasonally classified inputs for the prediction of daily groundwater levels: NARX networks vs support vector machines. Environ Model Assess 24(2):223–234. https://doi.org/10.1007/s10666-018-9639-x

Haykin S (2004) Neural networks: a Comprehensive Foundation. Prentice Hall, New Jersey

Iqbal M, Naeem UA, Ahmad A, Ghani U, Farid T (2020) Relating groundwater levels with meteorological parameters using ANN technique. Measurement 166:108163. https://doi.org/10.1016/j.measurement.2020.108163

Ivakhnenko AG (1968) The group method of data of handling; a rival of the method of stochastic approximation. Soviet Automatic Control 13:43–55

Jaafari A, Panahi M, Mafi-Gholami D, Rahmati O, Shahabi H, Shirzadi A, Pradhan B (2022) Appl Soft Comput 116:108254. https://doi.org/10.1016/j.asoc.2021.108254. Swarm intelligence optimization of the group method of data handling using the cuckoo search and whale optimization algorithms to model and predict landslides.

Jang JS (1993) ANFIS: adaptive-network-based fuzzy inference system. IEEE Trans Syst man Cybernetics 23(3):665–685. https://doi.org/10.1109/21.256541

Kamali MZ, Davoodi S, Ghorbani H, Wood DA, Mohamadian N, Lajmorak S, Band SS (2022) Permeability prediction of heterogeneous carbonate gas condensate reservoirs applying group method of data handling. Mar Pet Geol 139:105597. https://doi.org/10.1016/j.marpetgeo.2022.105597

Khedri A, Kalantari N, Vadiati M (2020) Comparison study of artificial intelligence method for short term groundwater level prediction in the northeast Gachsaran unconfined aquifer. Water Supply 20(3):909–921. https://doi.org/10.2166/ws.2020.015

Khodakhah H, Aghelpour P, Hamedi Z (2022) Comparing linear and nonlinear data-driven approaches in monthly river flow prediction, based on the models SARIMA, LSSVM, ANFIS, and GMDH. Environ Sci Pollut Res 29(15):21935–21954. https://doi.org/10.1007/s11356-020-10543-3

Kiyani V, Esmaili A, Alijani F, Samani S, Vasić L (2022) Investigation of drainage structures in the karst aquifer system through turbidity anomaly, hydrological, geochemical and stable isotope analysis (Kiyan springs, western Iran). Environ Earth Sci 81(22):517

Koza JRGP (1992) On the programming of computers by means of natural selection. Genetic programming

Kumar M, Kar IN (2009) Nonlinear HVAC computations using least square support vector machines. Energy Conv Manag 50(6):1411–1418. https://doi.org/10.1016/j.enconman.2009.03.009

Lemke F (1997) Knowledge extraction from data using self-organizing modeling technologies. In Proceedings of the SEAM'97 Conference

Li D, Armaghani DJ, Zhou J, Lai SH, Hasanipanah M (2020) A GMDH predictive model to predict rock material strength using three non-destructive tests. J Nondestr Eval 39(4):1–14. https://doi.org/10.1007/s10921-020-00725-x

Lin L, Li S, Sun S, Yuan Y, Yang M (2020) A novel efcient model for gas compressibility factor based on GMDH network.

Flow Meas Instrum 71:101677. https://doi.org/10.1016/j.flowmeasinst.2019.101677

Mathworks M (2014) Fuzzy logic toolbox. User's Guide, The Mathworks, Massachusetts

McGarry KJ, Wermter S, MacIntyre J (1999) Knowledge extraction from radial basis function networks and multilayer perceptrons. In IJCNN'99. International Joint Conference on Neural Networks. Proceedings (Cat. No. 99CH36339) (Vol. 4, pp. 2494–2497. IEEE https://doi.org/10.1109/IJCNN.1999.833464

Mehdizadeh S, Behmanesh J, Khalili K (2017) Application of gene expression programming to predict daily dew point temperature. Appl Therm Eng 112:1097–1107. https://doi.org/10.1016/j.applthermaleng.2016.10.181

Miraki S, Zanganeh SH, Chapi K, Singh VP, Shirzadi A, Shahabi H, Pham BT (2019) Mapping groundwater potential using a novel hybrid intelligence approach. Water Resour Manage 33(1):281–302. https://doi.org/10.1007/s11269-018-2102-6

Moghaddam HK, Milan SG, Kayhomayoon Z, Azar NA (2021) The prediction of aquifer groundwater level based on spatial clustering approach using machine learning. Environ Monit Assess 193(4):1–20. https://doi.org/10.1007/s10661-021-08961-y

Mohajerani M, Dokhanian F, Estaji H, Boer D, Norouzi M (2024) Geospatial distribution of qanats in middle eastern countries: potential for sustainable groundwater system. J Arid Environ 222:105170

Molle F, Mamanpoush A, Miranzadeh M (2004) Robbing Yadullah's water to irrigate Saeid's garden: Hydrology and water rights in a village of central Iran (Vol. 80). IWMI

Moravej M, Amani P, Hosseini-Moghari SM (2020) Groundwater level simulation and forecasting using interior search algorithm-least square support vector regression (ISA-LSSVR). Groundw Sustainable Dev 11:100447. https://doi.org/10.1016/j.gsd.2020.100447

Moriasi DN, Gitau MW, Pai N, Daggupati P (2015) Hydrologic and water quality models: performance measures and evaluation criteria. Trans ASABE 58(6):1763–1785. https://doi.org/10.13031/trans.58.10715

Mozaffari S, Javadi S, Moghaddam HK, Randhir TO (2022) Forecasting Groundwater levels using a hybrid of support Vector Regression and particle swarm optimization. Water Resour Manage 1–18. https://doi.org/10.1007/s11269-022-03118-z

Mueller JA, Ivachnenko AG, Lemke F (1998) GMDH algorithms for complex systems modelling. Math Comp Model Dyn Sys 4(4):275–316.

Mulashani AK, Shen C, Nkurlu BM, Mkono CN, Kawamala M (2022) Enhanced group method of data handling (GMDH) for permeability prediction based on the modified Levenberg Marquardt technique from well log data. Energy 239:121915. https://doi.org/10.1016/j.energy.2021.121915

Nadiri AA, Habibi I, Gharekhani M, Sadeghfam S, Barzegar R, Karimzadeh S (2022) Introducing dynamic land subsidence index based on the ALPRIFT framework using artificial intelligence techniques. Earth Sci Inf 1–15. https://doi.org/10.1007/s12145-021-00760-w

Naghibi SA, Pourghasemi HR, Abbaspour K (2018) A comparison between ten advanced and soft computing models for groundwater Qanat potential assessment in Iran using R and GIS. Theoret Appl Climatol 131(3):967–984. https://doi.org/10.1007/s00704-016-2022-4

Najafabadipour A, Kamali G, Nezamabadi-Pour H (2022) Application of Artificial Intelligence techniques for the determination of Groundwater Level using spatio–temporal parameters. ACS Omega 7(12):10751–10764. https://doi.org/10.1021/acsomega.2c00536

Nariman-Zadeh N, Darvizeh A, Darvizeh M, Gharababaei H (2002) Modelling of explosive cutting process of plates using

GMDH-type neural network and singular value decomposition. J Mater Process Technol 128(1–3):80–87. https://doi.org/10.1016/S0924-0136(02)00264-9

Nasiri F, Mafakheri MS (2015) Qanat water supply systems: a revisit of sustainability perspectives. Environ Syst Res 4(1):1–5. https://doi.org/10.1186/s40068-015-0039-9

Patel MB, Patel JN, Bhilota UM (2022) Comprehensive Modelling of ANN. In Research Anthology on Artificial neural network applications. IGI Global 31–40. https://doi.org/10.4018/978-1-6684-2408-7.ch002

Pham QB, Kumar M, Di Nunno F, Elbeltagi A, Granata F, Islam ARM, Anh DT (2022) Groundwater level prediction using machine learning algorithms in a drought-prone area. Neural Comput Appl 1–23. https://doi.org/10.1007/s00521-022-07009-7

Platt JC (1999) Fast training of support vector machines using sequential minimal optimization, advances in kernel methods. Support Vector Learn 185–208. https://doi.org/10.1109/ISKE.2008.4731075

Poursaeid M, Poursaeid AH, Shabanlou S (2022) A comparative study of Artificial Intelligence models and A Statistical Method for Groundwater Level Prediction. Water Resour Manage 1–21. https://doi.org/10.1007/s11269-022-03070-y

Sahoo S, Jha MK (2013) Groundwater-level prediction using multiple linear regression and artificial neural network techniques: a comparative assessment. Hydrogeol J 21(8):1865–1887. https://doi.org/10.1007/s10040-013-1029-5

Samani S (2024) Unraveling aquifer dynamics: Time series evaluation for informed groundwater management. Groundw Sustainable Dev 25:101174

Samani S, Boustani F, Hojati MH (2013) Screen for heavy metals from groundwater samples from industrialized zones in Marvdasht, Kharameh and Zarghan plains, Shiraz, Iran. World Appl Sci J 22(3):380–388

Samani S, Vadiati M, Azizi F, Zamani E, Kisi O (2022) Groundwater level simulation using soft computing methods with emphasis on major meteorological components. Water Resour Manage 36(10):3627–3647

Samani S, Vadiati M, Delkash M, Bonakdari H (2023) A hybrid wavelet–machine learning model for qanat water flow prediction. Acta Geophys 71(4):1895–1913

Samani S, Vadiati M, Nejatijahromi Z, Etebari B, Kisi O (2023b) Groundwater level response identification by hybrid wavelet–machine learning conjunction models using meteorological data. Environ Sci Pollut Res 30(9):22863–22884

Samantaray S, Sahoo A (2023) Prediction of flow discharge in Mahanadi River Basin, India, based on novel hybrid SVM approaches. Environment, Development and Sustainability, pp 1–25

Samantaray S, Biswakalyani C, Singh DK, Sahoo A, Satapathy P, D (2022) Prediction of groundwater fluctuation based on hybrid ANFIS-GWO approach in arid Watershed, India. Soft Comput 1–23. https://doi.org/10.1007/s00500-022-07097-6

Samantaray S, Sahoo A, Agnihotri A (2023a) Prediction of flood discharge using hybrid PSO-SVM algorithm in Barak River Basin. MethodsX 10:102060

Samantaray S, Sahoo P, Sahoo A, Satapathy DP (2023b) Flood discharge prediction using improved ANFIS model combined with hybrid particle swarm optimisation and slime mould algorithm. Environ Sci Pollut Res 30(35):83845–83872

Sanikhani H, Kisi O (2012) River flow estimation and forecasting by using two different adaptive neuro-fuzzy approaches. Water Resour Manage 26(6):1715–1729. https://doi.org/10.1007/s11269-012-9982-7

Sedghi MM, Zhan H (2020) Semi-analytical solutions of discharge variation of a Qanat in an unconfined aquifer subjected to general areal recharge and nearby pumping well discharge. J Hydrol 584:124691. https://doi.org/10.1016/j.jhydrol.2020.124691

Sedghi MM, Zhan H (2024) Discharge variations of qanat near an ephemeral stream. J Hydrol, 131367

Sreelakshmi S, Shaji E (2022) Landslide identification using machine learning techniques: review, motivation, and future prospects. Earth Sci Inf 15(4):2063–2090

Sridharam S, Sahoo A, Samantaray S, Ghose DK (2021) Estimation of water table depth using wavelet-ANFIS: a case study. Communication Software and Networks. Springer, Singapore, pp 747–754. https://doi.org/10.1007/978-981-15-5397-4_76

Sun J, Hu L, Li D, Sun K, Yang Z (2022) Data-driven models for accurate groundwater level prediction and their practical significance in groundwater management. J Hydrol 608:127630. https://doi.org/10.1016/j.jhydrol.2022.127630

Suryanarayana C, Sudheer C, Mahammood V, Panigrahi BK (2014) An integrated wavelet-support vector machine for groundwater level prediction in Visakhapatnam, India. Neurocomputing 145:324–335. https://doi.org/10.1016/j.neucom.2014.05.026

Suykens JA, Vandewalle J (1999) Least squares support vector machine classifiers. Neural Process Lett 9(3):293–300. https://doi.org/10.1023/A:1018628609742

Tao H, Hameed MM, Marhoon HA, Zounemat-Kermani M, Salim H, Sungwon K, Yaseen ZM (2022) Groundwater Level Prediction using machine learning models: a Comprehensive Review. Neurocomputing. https://doi.org/10.1016/j.neucom.2022.03.014

Tao H, Abba SI, Al-Areeq AM, Tangang F, Samantaray S, Sahoo A, Yaseen ZM (2024) Hybridized artificial intelligence models with nature-inspired algorithms for river flow modeling: a comprehensive review, assessment, and possible future research directions. Eng Appl Artif Intell 129:107559

Tayebi HA, Ghanei M, Aghajani K, Zohrevandi M (2019) Modeling of reactive orange 16 dye removal from aqueous media by mesoporous silica/crosslinked polymer hybrid using RBF, MLP and GMDH neural network models. J Mol Struct 1178:514–523. https://doi.org/10.1016/j.molstruc.2018.10.040

Tijani IA, Zayed T (2022) Gene expression programming based mathematical modeling for leak detection of water distribution networks. Measurement 188:110611. https://doi.org/10.1016/j.measurement.2021.110611

Vapnik V (1998) Statistical learning theory. John wiley&sons. Inc., New York, p 1

Wee WJ, Zaini NAB, Ahmed AN, El-Shafie A (2021) A review of models for water level forecasting based on machine learning. Earth Sci Inf 14:1707–1728

Yazdi AAS, Khaneiki ML (2016) Qanat knowledge: construction and maintenance. Springer