Check for updates

Predicting monthly evaporation from dam reservoirs using LS-SVR and ANFIS optimized by Harris hawks optimization algorithm

Naser Arya Azar · Sami Ghordoyee Milan [ID] ·
Zahra Kayhomayoon

Abstract Evaporation is a crucial factor in hydrological studies; its precise measurement has always been challenging due to the costly recording tolls. Therefore, machine learning models that can give reliable predictive results with the least information available have been recommended for evaporation prediction. This study was conducted in the central of Iran using the data related to the Doroudzan dam. Several hydrological and meteorological variables, including inflow and outflow of the reservoir, lake area behind the dam, temperature, overflow from the reservoir, precipitation, and evaporation at the previous month, were considered input data to predict the evaporation at the current month. Monthly data from October 1999 to September 2020 were used during the modeling. First, the single adaptive neuro-fuzzy inference system (ANFIS) and least-squares support vector regression (LS-SVR) models were evaluated for predicting the amount of evaporation using different scenarios defined based on the different combinations of input variables. The results showed that LS-SVR with RMSE = 2.77, MAPE = 2.48, and NSE = 0.93 provided a better prediction than ANFIS. Second, the Harris hawks optimization (HHO) algorithm was used to optimize the parameters of ANFIS to check for the possibility of performance improvement. The hybrid ANFIS-HHO model predicted the evaporation with RMSE = 2.35, MAPE = 1.55, and NSE = 0.95, respectively. The Taylor's diagram also demonstrated the superior performance of the hybrid ANFIS-HHO model than the LS-SVR and ANFIS models. The best scenario for all three models included all input variables but the area behind the dam into the models. The methodology proposed in this study is useful for predicting the evaporation from dam reservoirs under the influence of various dam variables.

N. Arya Azar
Department of Water Engineering, Faculty of Agriculture, University of Tabriz, Tabriz, Iran
e-mail: naseraryaazar92@gmail.com

S. Ghordoyee Milan (✉)
Department of Irrigation and Drainage Engineering, Aburaihan Campus, University of Tehran, Tehran, Iran
e-mail: s.milan@ut.ac.ir

Z. Kayhomayoon
Department of Geology, Payame Noor University, Tehran, Iran
e-mail: Zkayhomayoon@pnu.ac.ir

Introduction

The importance of water and how it affects human life is evident, as it is impossible to survive without water. Therefore, proper water resources' management is essential and plays a crucial role in the future of human beings (Orimoloye et al., 2020; Weng et al.,

2021). It is necessary to have sufficient knowledge and understanding of all the factors involved in the resources' development and limitation (Orimoloye et al., 2021; Quan et al., 2021). Evaporation is one of the most critical factors that play an essential role in the hydrological cycle. Evaporation is an essential hydrological variable in the study, control, and management of water resources (Friedrich et al., 2018).

Increased evaporation is a remarkable indicator of global warming (Chen et al., 2018; Limjirakan & Limsakul, 2012). Monitoring changes in evaporation is of great importance for water resources' monitoring and management (Kim et al., 2013; Wang 2020). Water losses due to evaporation significantly affect the water budget of reservoirs and lakes which in turn can remarkably decrease the water level. Consequently, water loss due to evaporation should be determined before designing irrigation systems and adopting water resource strategies (Allawi et al., 2019). Reliable evaporation prediction is a critical aspect of the hydrological considerations in water resources' management, water balance, and water use improvement. The use of previous information available for these variables makes it possible to predict future developments that are a key factor in the planning, design, and management of water resources (Owolabi et al., 2020). Therefore, to accurately predict the amount of evaporation, a relatively long period of previous data is required along with hydrological and meteorological information. This information should be variable with respect to time because if the values of a variable are constant over time, it will not affect evaporation changes.

The evaporation rate is affected by various climatic variables, such as temperature, lake surface area, and precipitation (Benzaghta et al., 2012). In other words, these factors create complex and nonlinear equations describing evaporation. The prediction of evaporation requires complicated nonlinear equations with several input variables (Sebbar et al., 2019). However, it is not practical and realistic to consider many physical variables and factors for predicting the evaporation rates (Rianna et al., 2018), although climatic variables and the inlet and outlet information of the dam are required to predict the evaporation from the reservoir dam correctly.

Direct and indirect methods, including water balance, energy balance, mass transfer, Penman, and evaporation pan, are used to predict evaporation (Wu et al., 2020). Among these methods, the evaporation pan method has been widely used due to its low cost and simple operation (Keshtegar et al., 2016). However, the installation and maintenance of the pan in some places are impossible, or daily reading of the evaporation rate is challenging (Kişi, 2006). In indirect methods, evaporation is estimated using meteorological data and energy volume and energy conservation relationships, which require calibration in regions with different climates. However, it has been proven that both these methods cannot provide reliable estimations of evaporation. Both methods' unsatisfactory performance has led water scientists to test other approaches for evaporation prediction (Quinn et al., 2018).

According to the literature, machine learning has been successfully applied for water resource problems, for example, rainfall, runoff, sedimentation, river flow, water level, water quality, and reservoir operations (Adnan et al., 2021; He, et al., 2018; Nhu et al., 2020). These methods are data-driven and do not require physical information from the study area. They identify patterns embedded in time series information and use these patterns to predict future scenarios. Recent studies have shown that these methods can achieve more accurate results than other models in hydrological applications (Arya Azar et al., 2021; Chu & Chang, 2009) and other fields of research (Jiang et al., 2017; He, 2020).

Machine learning methods have also been successfully used in evaporation studies (Ghorbani et al., 2018; Allawi et al., 2019; Wu et al., 2019). Wu et al. (2020) used machine learning models to predict monthly evaporation from the evaporation pan and reported on the acceptable performance of machine learning models in predicting monthly evaporation. Antonopoulos and Antonopoulos (2017) used artificial neural networks (ANN) with experimental methods for predicting daily evaporation data and reported that the ANN model provided better results in evaporation prediction. Goyal et al. (2014) utilized ANN, least-squares support vector regression (LS-SVR), and fuzzy inference system (FIS) to predict the daily evaporation of the pan in subtropical climates.

In previous studies, evaporation from reservoir dams has rarely been discussed. Since the amount of evaporation from the surface of the reservoirs is one of the essential parameters of water balance, its correct prediction is essential in hydrological studies.

Therefore, this study was aimed to predict the amount of evaporation from the surface of dam reservoirs using two machine learning models and also using evolutionary algorithms. Due to the importance of evaporation and studies on predicting the evaporation amounts, the performance of LS-SVR and ANFIS was evaluated in evaporation prediction. Then, to improve the ANFIS prediction performance, Harris hawks optimization (HHO; Heidari et al., 2019) was

considered for optimizing the parameters of ANFIS. Afterward, the developed ANFIS-HHO hybrid model was utilized for predicting the monthly evaporation from the dam reservoir. Different scenarios of input variables were developed and incorporated into each model. The results of the models and scenarios were analyzed, and the best model with the most appropriate scenario was selected and proposed for predicting the evaporation from reservoir dams.
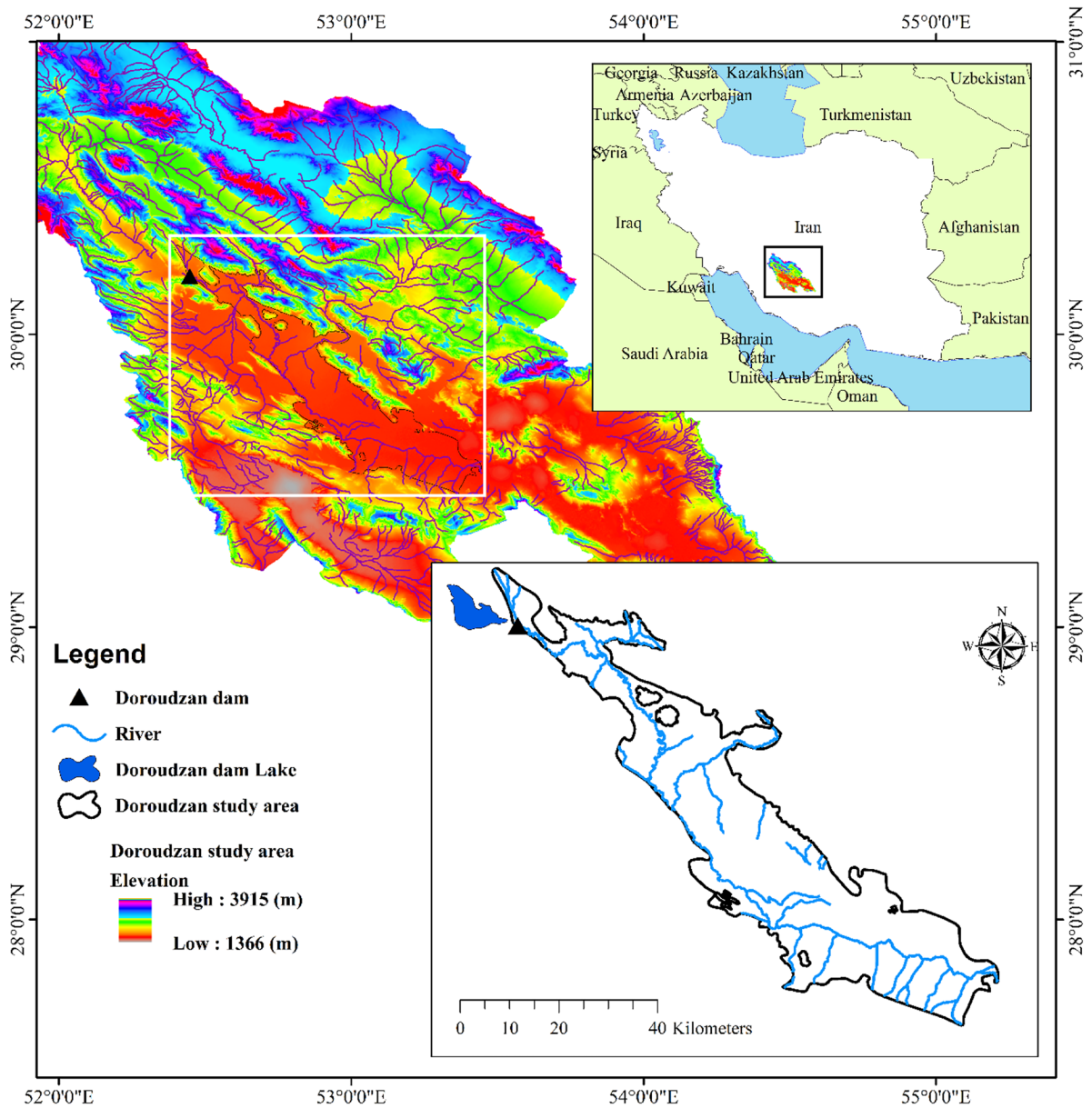

Fig. 1 The location of the study area in Iran

Study area and the data used

The Doroudzan dam (29° 50′−30° 15′ N, 51° 53′−52° 22′ E) is located in Tasht-e Bakhtegan watershed, 100 km northwest of Shiraz, central Iran, on the Kor River. By supplying about 760 MCM/year of water, this dam provides agricultural water of ca. 42,000 ha of Ramjerd and 34,000 ha of Korbal and Marvdasht. The area reported for the watershed is 4,116 km². Figure 1 shows the geographical location of the study area.

Several variables, including temperature (T), inflow to the dam reservoir (Q_{in}), the outflow from the dam reservoir (Q_{out}), overflow from the dam reservoir (OF), lake area behind the dam (A), precipitation (P), and evaporation at previous month ($EVO_{(n-1)}$), were used to predict the monthly evaporation (EVO). The statistical characteristics of these variables during the study period are shown in Table 1. The evaporation varied in the range of 0 to 74.5 mm per month. The lowest evaporation amounts were in the cold months of the year: January and February. Moreover, the highest amounts of evaporation were recorded in the hot months of the year. The amount of monthly precipitation during the study period varied from 0 to 730.5 mm per month. Moreover, when the amount of rainfall during the day increased, the amount of inflow to the reservoir naturally increased. The maximum temperature was about 11.93 °C, while its average value was about 5.25 °C.

A, lake area behind the dam; OF, overflow from the dam reservoir; T, temperature; Q_{out}, the outflow from the dam reservoir; Q_{in}, inflow to the dam reservoir; P, precipitation; EVO, evaporation.

Methodology

Since this study was aimed to provide a reliable and efficient model for predicting evaporation from dam reservoirs, the hydrological and meteorological variables affecting the evaporation were determined. These variables included lake area behind the dam, precipitation, inflow to the reservoir, outflow from the reservoir, temperature, and overflow. Then, various input scenarios of variables were defined, which were then implemented by machine learning techniques. The HHO evolutionary algorithm was used to optimize the ANFIS parameters to improve its prediction performance. According to the literature, the HHO algorithm has unique features that can significantly improve the ANFIS model. In this structure, the objective function is defined as minimizing the error of the values predicted by the model. The performance of the models and scenarios was investigated using error evaluation criteria in the form of statistical and graphical relationships (i.e., RMSE, MAPE, NSE, Taylor's diagram, and scatterplots). Finally, the most appropriate predictive model with its appropriate scenario was proposed (Fig. 2). In the following, the models used in this study are described in detail.

Least-squares support vector regression (LS-SVR)

Conventional SVR often fails to provide optimum solution optimization problems, which can lower the performance of the machine. Therefore, LS-SVR is recommended for solving complex problems since it exerts lower computational complexity compared to SVR, resulting in more desirable performance (Arya Azar et al., 2021; Goyal et al., 2014).

Given a set of training data such as $\{x_k, y_k\}_{K=1}^N$, whose input and output data include $x_k \in R^N$ and $y_k \in R$, respectively, Eq. (1) shows the nonlinear regression function in the initial weighting (Suykens and Vandewalle, 1999)

$$y((x)) = W^T \varphi((x)) + b \tag{1}$$

where T, b, and W are the weight, regression bias, and transpose operator, respectively. $\varphi(x)$ maps the

Table 1 Statistical characteristics of the variables used in this study

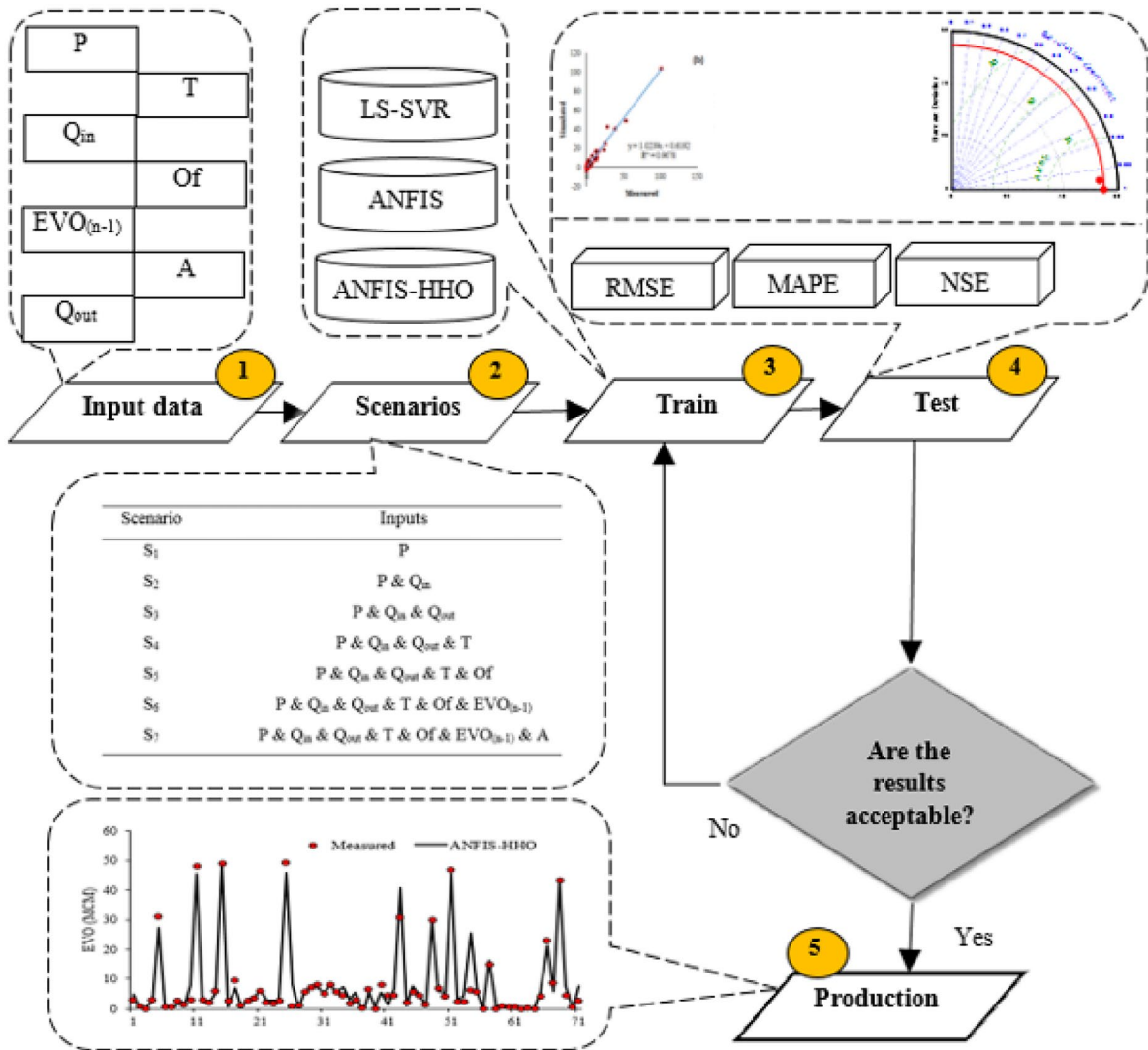|	A (km²)	OF (MCM)	T (°C)	Q_{out} (MCM)	Q_{in} (MCM)	P (mm)	EVO (MCM)
Minimum	26.2	0.00	− 1.85	0.28	0.160	0.0	0.00
Maximum	56.0	292	11.93	4344	1604	730.5	74.5
Average	35.7	5.06	5.25	144	100.8	62.91	8.14
Standard deviation	7.14	29.6	3.85	401	211.8	123.5	12.21

Fig. 2 The flowchart of this study

inputs in the feature space with high dimensions. This nonlinear regression can be solved by optimizing Eq. (2).

$$\min j((w,\ e)) = \frac{1}{2}W^2 W + \frac{1}{2}\gamma \sum_{k=1}^{N} e_k^2 \qquad (2)$$

Subject to:

$$y_k = W^T \varphi(x) + b + e_k, k = 1, N \qquad (3)$$

where γ is the regulator parameter for the error e. γ always controls the approximation function, so

the larger the γ value, the higher the error. Solving this equation using the Lagrangian form of the main objective function:

$$L(w,b,e,a) = j(w,e) - \sum_{i=1}^{N} \alpha_i \{ W^T \varphi(x_k) + b + e_k - y_k \} \qquad (4)$$

where α_i is the Lagrangian coefficient. Based on the Karush–Kuhn–Tucker condition, the LS-SVR model is written for the approximation function as Eq. (5).

$$y(x) = \sum_{k=1}^{N} \alpha_k K(x, x_k) + b \tag{5}$$

where $K(x, x_k)$ is called kernel function. In this study, the Gaussian function (Eq. (6)) was used.

$$K(x, x_k) = exp(-\frac{||x - x_k||^2}{\sigma^2}) \tag{6}$$

Adaptive neuro-fuzzy inference system (ANFIS)

Jang (1993) developed the ANFIS model for the first time by combining ANN and fuzzy logic. ANFIS does not have ANN and FIS limitations, such as overfitting and sensitivity to the definition of membership functions, to perform better in prediction problems. The most common method for the training of ANFIS is the Sugeno-type FIS, which uses a robust learning algorithm to determine the model's parameters (Asefpour Vakilian & Massah, 2018). ANFIS architecture generally includes five layers. In layer 1, the generalized Gaussian membership function μ produces a new output Out_{1i} from the inputs x and y (Eq. (7)).

$$\begin{aligned} Out_{1i} &= \mu A_i(x), i = 1, 2 \\ Out_{1i} &= \mu B_{i-2}(y), i = 3, 4 \end{aligned} \tag{7}$$

Where

$$\mu(x) = e^{-\left(x - \frac{P_i}{\sigma_i}\right)^2} \tag{8}$$

and A_i and B_i are the membership values of μ, while P_i and σ_i are the equation parameters. The output of each node is obtained in the second layer using Eq. (9)

$$Out_{2i} = \mu A(x) * \mu B_{i-1} \tag{9}$$

Then, the output of layer 2 is normalized in layer 3 (Eq. (10))

$$Out_{3i} = \overline{W}_i = \frac{\omega_i}{\sum_{i=1}^{2} \omega_i} \tag{10}$$

The output is then used in a linear combination equation

$$Out_{4,1} = \overline{w}_i f_i = \overline{w}_i(p_i x + q_i y + r_i) \tag{11}$$

where p, q, and r are parameters defined for the ith node. The model's output is obtained using Eq. (12).

$$Out_{4,i} = \sum_i \overline{w}_i f_i \tag{12}$$

Harris hawks optimization (HHO)

Introduced by Heidari et al. (2019), the HHO algorithm is inspired by nature and how rabbits are hunted by Harris hawks. This algorithm involves two stages of soft and hard besieges of the rabbit. In the soft besiege, the rabbit still has enough energy and tries to escape with random misleading jumps. Harris hawks gently surround it to make the rabbit more tired. However, in the hard besiege, the prey is very tired and has little energy for escape. Finally, the hawks hardly encircle the rabbit for performing a surprise pounce.

In this algorithm, the Harris hawks move randomly to find prey. Their position is mathematically expressed as:

$$X(t+1) = \begin{cases} X_{rand}(t) - r_1|X_{rand}(t) - 2r_2 X(t)| & q \geq 0.5 \\ (X_{rabbit}(t) - X_m(t)) - r_3(LB + r_4(UB - LB)) & q < 0.5 \end{cases} \tag{13}$$

where $X(t)$ and $X(t+1)$ denote the position of hawks at iterations t and $t+1$, respectively; $X_{rabbit}(t)$ is the rabbit's position; r_1, r_2, r_3, r_4, and q are random numbers, being updated in each iteration; UB and LB are the lower and upper limits of variables; $X_{rand}(t)$ is the position of a hawk randomly selected from the population; and X_m is the average position of the population, which is obtained using Eq. (14).

$$X_m(t) = \frac{1}{N} \sum_{i=1}^{N} X_i(t) \tag{14}$$

where N is the total number of hawks and $X_i(t)$ is the position of each hawk in iteration t. The prey's energy decreases during the escape (Eq. (15))

$$E = 2E_0(1 - \frac{t}{T}) \tag{15}$$

where T is the maximum iteration number, E is the prey's energy, and E_0 is the initial energy. The E parameter is utilized for enabling the algorithm to use soft and hard besiege processes to trap the prey. Soft and hard besieging occur when $|E| \geq 0.5$ and $|E| < 0.5$, respectively.

When $|E| \geq 0.5$, although the prey performs some random misleading jumps since it still has enough

energy, it finally cannot. The hawks encircle it softly to make the rabbit more exhausted and then perform the surprise pounce (Eqs. (16) and (17)).

$$X(t + 1) = \Delta X(t) - E|JX_{\text{rabbit}}(t) - X(t)| \quad (16)$$

$$\Delta X(t) = X_{\text{rabbit}}(t) - X(t) \quad (17)$$

where $\Delta X(t)$ is the difference between the prey's position and the current position in iteration t, and J is a coefficient representing the strength of the prey's random jumps. When $|E| < 0.5$, the rabbit has low escaping energy since it is exhausted, and at this time, the surprise pounce is performed by the hawks. Equation (18) shows the updates of current positions in hard besiege.

$$X(t + 1) = X_{\text{rabbit}}(t) - E|\Delta X(t)| \quad (18)$$

Input scenarios

LS-SVR, ANFIS, and ANFIS-HHO were used in this study for predicting monthly evaporation from the Doroudzan reservoir dam. For this purpose, various scenarios with different combinations of effective variables were developed. The correlation coefficient of each variable with the output ($EVO_{(n-1)}$) is listed in Table 2. The variable with the highest correlation coefficient with the output was introduced as the first scenario. The second scenario was developed based on two variables that achieved the highest correlation coefficients. The rest of the scenarios were defined using other variables such that the S_7 included all input variables. The scenarios defined in this study to predict the evaporation from the dam are listed in Table 3. Each of these scenarios was implemented by each model, and its results were evaluated by the performance evaluation criteria.

In examining the correlation coefficients between the inputs and output, it is observed that evaporation at the previous month had a lower correlation than other parameters such as temperature, precipitation, and inflows and outflows. This shows that in predicting evaporation from dam reservoirs, parameters such as inflow and outflow of the dam can be more effective

Table 3 Scenarios defined in this study to predict the evaporation

Scenario	Inputs	Output
S_1	P	EVO
S_2	P & Q_{in}	
S_3	P & Q_{in} & Q_{out}	
S_4	P & Q_{in} & Q_{out} & T	
S_5	P & Q_{in} & Q_{out} & T & Of	
S_6	P & Q_{in} & Q_{out} & T & Of & $EVO_{(n-1)}$	
S_7	P & Q_{in} & Q_{out} & T & Of & $EVO_{(n-1)}$ & A	

than evaporation values in previous months. The evaporation had the highest correlation with precipitation, inflows and outflows of the dam, and temperature, respectively, and the lowest correlation with the reservoir surface area.

Performance evaluation criteria

The dataset was randomly divided into two groups: 70% of data were considered for model training and the remaining 30% were used for the test. Root mean square error (RMSE), mean absolute percentage error (MAPE), Nash Sutcliffe Index (NSE), and coefficient of determination (R^2) were considered for evaluating the scenarios and machine learning methods (Hua et al., 2021; Weng et al., 2021).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (x_o - x_p)^2}{n}} \quad (19)$$

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^{n} |\frac{x_o - x_p}{x_o}| \quad (20)$$

$$NSE = 1 - \frac{\sum_{i=1}^{n} (O_i - S_i)^2}{\sum_{i=1}^{n} (O_i - \overline{O})^2} \quad (21)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (x_p - x_o)^2}{\sum_{i=1}^{n} (x_o - \overline{x}_o)^2} \quad (22)$$

Table 2 The correlation coefficients between output and inputs

	$EVO_{(n-1)}$	P	Q_{in}	Q_{out}	T	Of	A
EVO	0.17	0.65	0.60	0.51	0.40	0.22	−0.160

where x_o is the observed (measured) value, x_p is the predicted value, and n is the number of samples. The lower RMSE and MAPE and higher NSE and R^2 values indicate better model performance.

Results and discussion

To obtain the most proper values for the parameters of each machine learning algorithm, it was necessary to run the algorithm several times with different parameter values. The results of this procedure are shown in Table 4 and indicate that the Gaussian function is the most appropriate fuzzy membership function for the ANFIS model. Since ANFIS uses the Sugeno-type method in its structure, the linear function was used for the model's output equation. Some ANFIS parameters cannot be obtained by the trial and error method and require robust optimization algorithms to obtain their optimized values. Therefore, HHO was used to optimize the ANFIS model. The population for HHO and its maximum iteration number were adjusted to 30 and 2000, respectively. The Gaussian kernel in LS-SVR had two parameters, namely, σ^2 and γ, the optimum values of which were obtained equal to 5.365 and 136.03, respectively.

Error evaluation criteria for the training and test data are presented in Table 5. A model is introduced as the superior model in which the RMSE and MAPE values are the lowest for both test and training data, while the NSE is the highest. Scenario S_6 that included all variables but the lake area behind the dam was identified as the most suitable scenario for all three models. The best performance was obtained using the ANFIS-HHO model with RMSE, MAPE, and NSE of 1.55, 2.35, and 0.95, respectively. Error evaluation criteria showed that ANFIS had the lowest accuracy among the models with RMSE, MAPE, and NSE of 3.85, 5.30, and 0.85, respectively. Therefore, the HHO algorithm improved the accuracy of ANFIS performance by RMSE, MAPE, and NSE of 1.55, 2.35, and 0.95, respectively. The performance of LS-SVR was slightly lower than that of ANFIS-HHO, with RMSE, MAPE, and NSE values equal to 2.48, 2.77, and 0.93, respectively. This shows that although LS-SVR had relatively small errors during the prediction, ANFIS-HHO was the most suitable model among the developed models in the prediction of evaporation of reservoir dams.

The first three scenarios (S_1 to S_3) had the lowest prediction accuracy in all three models, which indicate that the monthly evaporation from the dam reservoir is so complex and nonlinear that it is not possible to predict its amount by having only the inflow and outflow of the dam and evaporation at the previous month. Therefore, it requires more information, for example, the input variables of scenario S_6, which

Table 4 Parameters and specifications of the models for the prediction of evaporation

Parameter	Value		
	ANFIS	ANFIS-HHO	LS-SVR
Fuzzy structure	Sugeno-type	-	-
Initial FIS for training	Genfis3	-	-
The type of membership functions	Gaussian	-	-
The membership function of output	Linear	-	-
Optimization method	Hybrid	-	-
Number of fuzzy rules	10	-	-
The maximum number of epochs	2000	-	-
Number of search agent	-	40	-
Iteration number	-	2500	-
β	-	1.5	-
Range partitions (weights and biases)	-	$[-3, +3]$	-
Population size	-	30	-
Kernel function	-	-	RBF
γ	-	-	1206.7
σ^2	-	-	1481.3

Table 5 Performance evaluation of the developed models. The best performances are shown in bold

Scenario	Model	MAPE		RMSE		NSE	
		Train	Test	Train	Test	Train	Test
S_1	ANFIS	3.80	8.62	5.52	10.2	0.68	0.32
	ANFIS-HHO	4.37	4.40	7.65	8.60	0.65	0.52
	LS-SVR	4.15	4.75	7.46	8.05	0.52	0.51
S_2	ANFIS	4.30	9.35	5.68	12.2	0.68	0.42
	ANFIS-HHO	4.97	6.85	7.33	9.95	0.53	0.48
	LS-SVR	4.48	4.26	7.79	5.98	0.61	0.63
S_3	ANFIS	3.20	7.80	4.31	12.3	0.85	0.63
	ANFIS-HHO	3.52	4.24	4.47	7.38	0.88	0.85
	LS-SVR	2.75	2.97	4.17	4.57	0.85	0.81
S_4	ANFIS	2.78	6.74	3.89	10.8	0.91	0.79
	ANFIS-HHO	2.60	2.94	4.25	4.63	0.91	0.88
	LS-SVR	1.77	3.10	3.10	5.57	0.98	0.85
S_5	ANFIS	2.10	4.38	3.25	6.41	0.90	0.82
	ANFIS-HHO	2.87	2.41	4.23	3.95	0.89	0.92
	LS-SVR	2.09	3.08	3.49	4.80	0.92	0.89
S_6	ANFIS	**1.05**	**3.85**	**2.18**	**5.30**	**0.99**	**0.85**
	ANFIS-HHO	**1.35**	**1.55**	**2.19**	**2.35**	**0.98**	**0.95**
	LS-SVR	**1.39**	**2.48**	**2.34**	**2.77**	**0.97**	**0.93**
S_7	ANFIS	2.12	3.66	3.34	5.35	0.94	0.78
	ANFIS-HHO	2.78	2.66	3.74	3.58	0.91	0.92
	LS-SVR	2.01	2.83	3.00	4.70	0.92	0.88

accurately predicted the evaporation. Therefore, in addition to the information on the dam's inflow and outflow, we need information about temperature, overflow, and evaporation in the previous month.

The scatterplots of the observed (measured) and predicted values (Fig. 3) showed that the ANFIS model had poor performance compared to other methods ($R^2 = 0.898$). In contrast, ANFIS-HHO achieved the highest performance than the other models ($R^2 = 0.959$). Moreover, data are close to the bisector line in the ANFIS-HHO model, which reveals its small prediction error.

For further evaluation, Taylor's diagram was used to investigate the correlation between the predicted and observed evaporation, as well as their standard deviations (Fig. 4). The correlation coefficient for all three models ranged from 0.95 to 0.99, indicating the efficiency of all three models for evaporation prediction. ANFIS-HHO had the highest correlation coefficient than the other two models. Furthermore, the root mean squares deviation (RMSD) for ANFIS-HHO and LS-SVR was ca. 2, while its value for the ANFIS model was slightly higher than 3. Although

all three models could predict evaporation, the closest results to the observed values using Taylor's diagram were obtained by the ANFIS-HHO model. Therefore, HHO was able to optimize the ANFIS parameters for increasing the model performance.

Figure 5 shows that ANFIS-HHO correctly detected the evaporation changes in almost all test steps. However, in several steps, such as 13, 28, and 65, the predicted values had significant errors when the ANFIS model was used for the prediction. In other words, ANFIS was unable to predict the minimum and maximum values of evaporation, and this result is probably achieved when the ANFIS model was trapped at local optimization points. The results also show that the changes in evaporation were detected correctly by the LS-SVR model. However, in some steps, the predicted values involved remarkable errors. The ANFIS-HHO model had the highest ability to predict the evaporation data; the minimum and maximum values are predicted with the lowest errors.

According to the results, to predict the evaporation properly, the use of meteorological data and some parameters related to the dam, such as its

Fig. 3 Scatter plots of the measured and predicted data
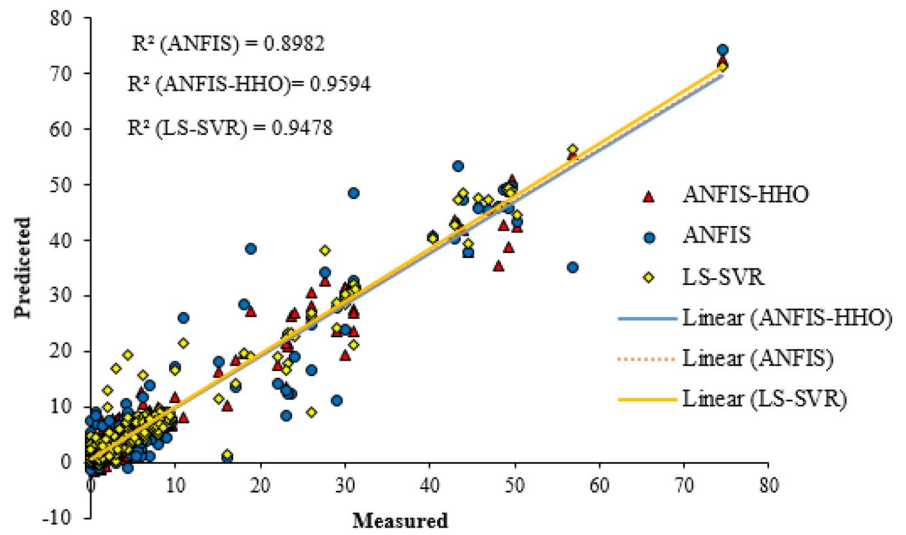
inflows and outflows and the area behind the dam, is required. Moreover, the correlation of the input parameters with the output showed that evaporation had a weak correlation with the evaporation at the previous month, indicating that only having various

delays of the target parameter could not result in a reliable prediction for future months. Therefore, to have an accurate prediction, information about temperature, precipitation, and dam inflows and outflows is required. The results also showed that the

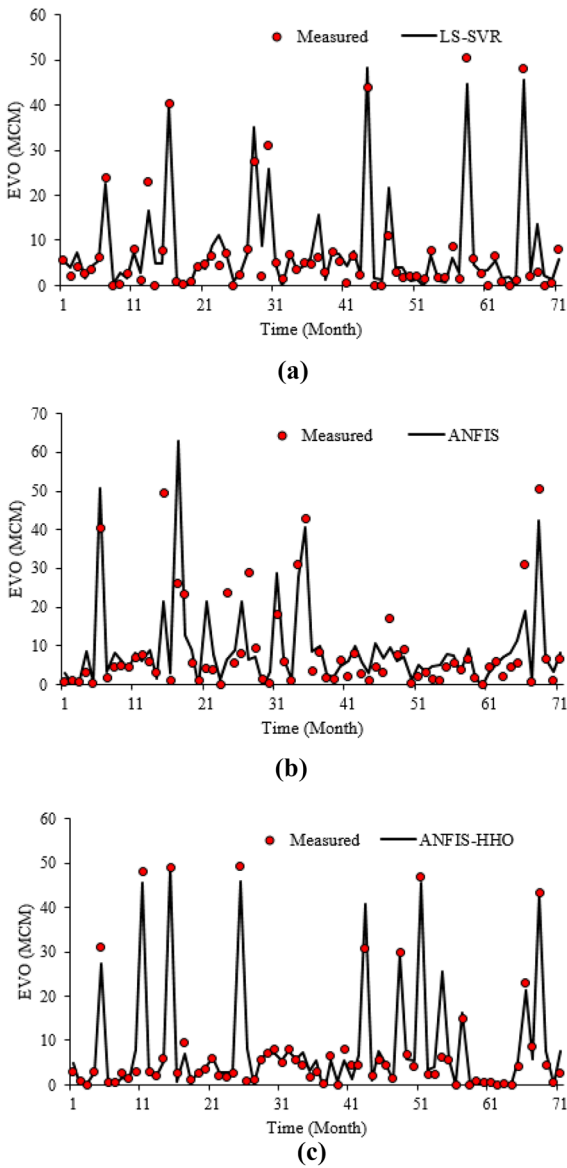**Fig. 4** Taylor's diagram for evaporation prediction
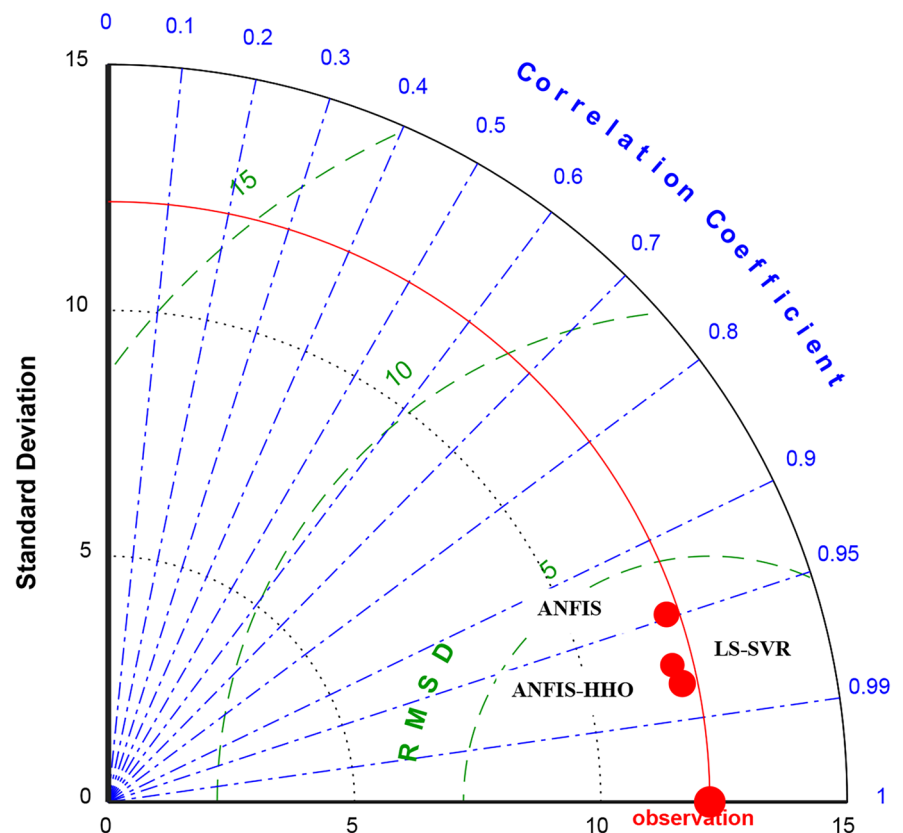
Fig. 5 Time series of observed and predicted values for evaporation using **a** LS-SVR, **b** ANFIS, and **c** ANFIS-HHO

presence of the water area behind the dam as an input parameter did not affect the prediction performance, so it could be omitted in the modeling. Of course, this result is slightly different from our understandings. We usually consider the lake surface area behind the dam as one of the effective parameters in evaporation losses. This is one of the disadvantages of the machine learning models that they do not consider the nature and type of data and

only prefer inputs with values being in line with the system output changes. Scenario S_6, including all input parameters but the surface area behind the dam, was the selected scenario for all three models, indicating the necessity of the participation of the parameters investigated in this study. Moreover, the results of Taylor's diagram, as well as scatter point, confirmed the results of RMSE, MAPE, and NSE error evaluation criteria in the promising performance of the ANFIS-HHO model.

In this study, for the first time, meteorological parameters along with dam inlet and outlet parameters were used to improve the prediction performance, which included a higher number of input variables than Allawi et al. (2021). In fact, this study tries to use the parameters of the dam balance to measure the amount of evaporation change in the future according to the changes of each input variable, which can help better manage the dam allocation. Therefore, in addition to the accurate prediction of evaporation, this study aimed to contribute meteorological variables and flow continuity parameters in the dam with appropriate accuracy to predict the evaporation amount in the future. This prediction can help us to make the right decision in the future in reducing evaporation or proper allocation of the dam. Developing various input scenarios allows researchers and authorities to consider different input information to predict evaporation depending on the status of each region. For example, in some areas, temperature information might not be available. In this case, we can use a scenario that does not include this parameter and has a relatively good performance, such as the third scenario introduced in this study. On the other hand, the use of meteorological parameters along with dam inlet and outlet variables provides a more realistic situation during the modeling, leading to a better prediction of evaporation.

In general, the performance of all three models was appropriate in predicting the amount of evaporation. Among the two single models, LS-SVR performed better than ANFIS, which was consistent with the results of Razavi et al. (2019) in estimating thermal conductivity enhancement and Bemani et al. (2019) in estimating the acid solvent solubility in supercritical CO_2 conditions. The use of the HHO algorithm improved the performance of ANFIS, which is consistent with the study of Arya Azar et al. (2021) in predicting the longitudinal dispersion coefficient of

the river, Milan et al. (2021) in predicting optimal groundwater withdrawal, and Shehabeldeen et al. (2019) in predicting the friction process of welding. Since various algorithms are proposed daily by researchers to solve optimization problems (Bo et al., 2021), the application of other algorithms is recommended for improving the performance of weak single models (e.g., ANFIS) in hydrological problems.

One of the strengths of using machine learning models is that they can predict evaporation without special knowledge of geology or meteorology. However, the results showed that an efficient prediction requires the participation of effective input variables such as temperature, precipitation, and inflow to the dam. Therefore, more information than one input parameter, such as evaporation at the previous month, is required to predict evaporation at the current month, which is in line with the findings of Allawi et al. (2021). Moreover, one of the advantages of using ANFIS and its hybrid models is considering the uncertainties in the input information, which did not exist in the LS-SVR model. On the other hand, although the HHO algorithm was able to improve the performance of ANFIS and the hybrid model developed had better efficiency than the LS-SVR model, the development of an LS-SVR model is much simpler than the ANFIS-HHO hybrid model. Therefore, in developing hybrid models based on metaheuristic optimization methods, in addition to improving the prediction performance, one should also pay attention to their structure and complexities. Hence, experts in the field of machine learning are required to develop such models since implementing the LS-SVR is much easier than the ANFIS-HHO.

The results showed that the performance of the models used strongly depends on their input variables. The variables with similar trends to the target parameter have a higher correlation coefficient with the output than other input variables and, therefore, are more important in determining the amount of output. This is a relatively fundamental weakness in the use of machine learning models because an input parameter might actually have a remarkable effect on evaporation but is not considered an important factor in machine learning models due to its trend of changes (e.g., area behind the dam in this study). Finally, it can be said that the trend of data is more valuable than the nature of the data for machine learning models.

Conclusions

The present study evaluated the performance of ANFIS and LS-SVR for the prediction of monthly evaporation from dam reservoirs. Seven scenarios that included different combinations of input variables were considered to evaluate the models' performance. LS-SVR performed better than the ANFIS model. To improve the ANFIS performance, the HHO algorithm optimized the ANFIS parameters. Among the input variables, precipitation, inflow to the dam, and temperature had the most significant effects on evaporation. The area of the lake behind the dam had the most negligible impact compared to other parameters. Two approaches were used to evaluate and select the appropriate model. In the first approach, error evaluation criteria (RMSE, MAPE, and NSE) were used to select the appropriate model and scenario, which showed that ANFIS-HHO is more accurate than the other two models. In the second approach, Taylor's diagram and scatterplots were used to compare the models graphically. Taylor's diagram reveals the correlation coefficient, standard deviation, and RMSD of predicted and observational data in the predictive models. Taylor's diagram showed that ANFIS-HHO resulted in the closest prediction values to the observational data. The introduced approach in this study can be used to predict and manage dams that have similar conditions to the dam investigated in this study.

In general, the results showed that the use of machine learning models to predict evaporation from reservoir dams provides satisfactory results that can be used in hydrological studies and management strategies. Furthermore, to have an accurate prediction, the information about the inflow and outflow of the dam, precipitation, and temperature was more effective than parameters such as evaporation at the previous month and the water area behind the dam. Due to the wide range of available machine learning models, it is recommended to evaluate these models to achieve the highest performance in predicting the evaporation of reservoirs in daily and monthly time steps. Considering climate change and investigating the daily prediction of evaporation from reservoir dams and including the amount of daily and monthly evaporation changes in the dam outflow planning program can be performed in future research.

Author contribution   N. Arya Azar: investigation, methodology, software, formal analysis, writing—original draft. S. Ghordoyee Milan: conceptualization, supervision data curation, software, visualization, writing—review and editing. Z. Kayhomayoon: validation, writing—review and editing.

Availability of data and materials   Not applicable.

Declarations

Ethics approval   All authors adhere to ethical approval.

Consent to participate   All authors agree.

Consent for publication   All authors agree.

Competing interests   The authors declare no competing interests.

References

Adnan, R. M., Jaafari, A., Mohanavelu, A., Kisi, O., & Elbeltagi, A. (2021). Novel ensemble forecasting of streamflow using locally weighted learning algorithm. *Sustainability, 13*(11), 5877.

Allawi, M. F., Aidan, I. A., & El-Shafie, A. (2021). Enhancing the performance of data-driven models for monthly reservoir evaporation prediction. *Environmental Science and Pollution Research, 28*(7), 8281–8295.

Allawi, M. F., Binti Othman, F., Afan, H. A., Ahmed, A. N., Hossain, M., Fai, C. M., & El-Shafie, A. (2019). Reservoir evaporation prediction modeling based on artificial intelligence methods. *Water, 11*(6), 1226.

Antonopoulos, V. Z., & Antonopoulos, A. V. (2017). Daily reference evapotranspiration estimates by artificial neural networks technique and empirical equations using limited input climate variables. *Computers and Electronics in Agriculture, 132*, 86–96.

Arya Azar, N., Milan, S. G., & Kayhomayoon, Z. (2021). The prediction of longitudinal dispersion coefficient in natural streams using LS-SVM and ANFIS optimized by Harris hawk optimization algorithm. *Journal of Contaminant Hydrology, 240*, 103781.

Asefpour Vakilian, K., & Massah, J. (2018). A fuzzy-based decision making software for enzymatic electrochemical nitrate biosensors. *Chemometrics and Intelligent Laboratory Systems, 177*, 55–63.

Bemani, A., Baghban, A., Shamshirband, S., Mosavi, A., Csiba, P., & Várkonyi-Kóczy, A. R. (2019). Applying ANN, ANFIS, and LSSVM models for estimation of acid solvent solubility in supercritical CO $ _2$. arXiv preprint arXiv:1912.05612.

Benzaghta, M. A., Mohammed, T. A., Ghazali, A. H., & Soom, M. A. M. (2012). Prediction of evaporation in tropical climate using artificial neural network and climate based models. *Scientific Research and Essays, 7*(36), 3133–3148.

Bo, W., Fang, Z. B., Wei, L. X., Cheng, Z. F., & Hua, Z. X. (2021). Malicious URLs detection based on a novel optimization algorithm. *IEICE TRANSACTIONS on Information and Systems, 104*(4), 513–516.

Chen, Y., He, L., Li, J., & Zhang, S. (2018). Multi-criteria design of shale-gas-water supply chains and production systems towards optimal life cycle economics and greenhouse gas emissions under uncertainty. *Computers & Chemical Engineering, 109*, 216–235.

Chu, H. J., & Chang, L. C. (2009). Application of optimal control and fuzzy theory for dynamic groundwater remediation design. *Water Resources Management, 23*(4), 647–660.

Friedrich, K., Grossman, R. L., Huntington, J., Blanken, P. D., Lenters, J., Holman, K. D., & Healey, N. C. (2018). Reservoir evaporation in the Western United States: Current science, challenges, and future needs. *Bulletin of the American Meteorological Society, 99*(1), 167–187.

Ghorbani, M. A., Deo, R. C., Yaseen, Z. M., Kashani, M. H., & Mohammadi, B. (2018). Pan evaporation prediction using a hybrid multilayer perceptron-firefly algorithm (MLP-FFA) model: Case study in North Iran. *Theoretical and Applied Climatology, 133*(3–4), 1119–1131.

Goyal, M. K., Bharti, B., Quilty, J., Adamowski, J., & Pandey, A. (2014). Modeling of daily pan evaporation in sub tropical climates using ANN, LS-SVR, fuzzy logic, and ANFIS. *Expert Systems with Applications, 41*(11), 5267–5276.

He, L., Chen, Y., & Li, J. (2018). A three-level framework for balancing the tradeoffs among the energy, water, and air-emission implications within the life-cycle shale gas supply chains. *Resources, Conservation and Recycling, 133*, 206–228.

He, Y., Dai, L., & Zhang, H. (2020). Multi-branch deep residual learning for clustering and beamforming in user-centric network. *IEEE Communications Letters, 24*(10), 2221–2225.

Heidari, A. A., Mirjalili, S., Faris, H., Aljarah, I., Mafarja, M., & Chen, H. (2019). Harris hawks optimization: Algorithm and applications. *Future Generation Computer Systems, 97*, 849–872.

Hua, L., Zhu, H., Shi, K., Zhong, S., Tang, Y., & Liu, Y. (2021). Novel finite-time reliable control design for memristor-based inertial neural networks with mixed time-varying delays. *IEEE Transactions on Circuits and Systems i: Regular Papers, 68*(4), 1599–1609.

Jang, J-SR. (1993) ANFIS: adaptive-network-based fuzzy inference system. I*EEE Transactions on Systems, Man, and Cybernetics, 23*(3), 665–685.

Jiang, Q., Shao, F., Lin, W., Gu, K., Jiang, G., & Sun, H. (2017). Optimizing multistage discriminative dictionaries for blind image quality assessment. *IEEE Transactions on Multimedia, 20*(8), 2035–2048.

Keshtegar, B., Piri, J., & Kisi, O. (2016). A nonlinear mathematical modeling of daily pan evaporation based on conjugate gradient method. *Computers and Electronics in Agriculture, 127*, 120–130.

Kim, S., Shiri, J., Kisi, O., & Singh, V. P. (2013). Estimating daily pan evaporation using different data-driven methods and lag-time patterns. *Water Resources Management, 27*(7), 2267–2286.

Kişi, Ö. (2006). Daily pan evaporation modelling using a neuro-fuzzy computing technique. *Journal of Hydrology, 329*(3–4), 636–646.

Limjirakan, S., & Limsakul, A. (2012). Trends in Thailand pan evaporation from 1970 to 2007. *Atmospheric Research, 108*, 122–127.

Milan, S. G., Roozbahani, A., Arya Azar, N., & Javadi, S. (2021). Development of adaptive neuro fuzzy inference system–Evolutionary algorithms hybrid models (ANFIS-EA) for prediction of optimal groundwater exploitation. *Journal of Hydrology*, *598*, 126258.

Nhu, V. H., Mohammadi, A., Shahabi, H., Shirzadi, A., Al-Ansari, N., Ahmad, B. B., & Nguyen, H. (2020). Monitoring and assessment of water level fluctuations of the Lake Urmia and its environmental consequences using multi-temporal landsat 7 ETM+ images. *International Journal of Environmental Research and Public Health, 17*(12), 4210.

Orimoloye, I. R., Belle, J. A., Olusola, A. O., Busayo, E. T., & Ololade, O. O. (2021). Spatial assessment of drought disasters, vulnerability, severity and water shortages: A potential drought disaster mitigation strategy. *Natural Hazards, 105*(3), 2735–2754.

Orimoloye, I. R., Kalumba, A. M., Mazinyo, S. P., & Nel, W. (2020). Geospatial analysis of wetland dynamics: Wetland depletion and biodiversity conservation of Isimangaliso Wetland, South Africa. *Journal of King Saud University-Science, 32*(1), 90–96.

Owolabi, S. T., Madi, K., Kalumba, A. M., & Orimoloye, I. R. (2020). A groundwater potential zone mapping approach for semi-arid environments using remote sensing (RS), geographic information system (GIS), and analytical hierarchical process (AHP) techniques: A case study of Buffalo catchment, Eastern Cape. *South Africa. Arabian Journal of Geosciences, 13*(22), 1–17.

Quan, Q., Gao, S., Shang, Y., & Wang, B. (2021). Assessment of the sustainability of Gymnocypris eckloni habitat under river damming in the source region of the Yellow River. *Science of The Total Environment, 778*, 146312.

Quinn, R., Parker, A., & Rushton, K. (2018). Evaporation from bare soil: Lysimeter experiments in sand dams interpreted using conceptual and numerical models. *Journal of Hydrology, 564*, 909–915.

Razavi, R., Sabaghmoghadam, A., Bemani, A., Baghban, A., Chau, K. W., & Salwana, E. (2019). Application of ANFIS and LSSVM strategies for estimating thermal conductivity enhancement of metal and metal oxide based nanofluids. *Engineering Applications of Computational Fluid Mechanics, 13*(1), 560–578.

Rianna, G., Reder, A., & Pagano, L. (2018). Estimating actual and potential bare soil evaporation from silty pyroclastic soils: Towards improved landslide prediction. *Journal of Hydrology, 562*, 193–209.

Sebbar, A., Heddam, S., & Djemili, L. (2019). Predicting daily pan evaporation (E pan) from dam reservoirs in the Mediterranean regions of Algeria: OPELM vs OSELM. *Environmental Processes, 6*(1), 309–319.

Shehabeldeen, T. A., Abd Elaziz, M., Elsheikh, A. H., & Zhou, J. (2019). Modeling of friction stir welding process using adaptive neuro-fuzzy inference system integrated with Harris hawks optimizer. *Journal of Materials Research and Technology, 8*(6), 5882–5892.

Suykens, J. A., & Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural Processing Letters*, *9*(3), 293–300.

Wang, Q., Wang, W., Zhong, Z., Wang, H., & Fu, Y. (2020). Variation in glomalin in soil profiles and its association with climatic conditions, shelterbelt characteristics, and soil properties in poplar shelterbelts of Northeast China. *Journal of Forestry Research, 31*(1), 279–290.

Weng, L., He, Y., Peng, J., Zheng, J., & Li, X. (2021). Deep cascading network architecture for robust automatic modulation classification. *Neurocomputing, 455*, 308–324.

Wu, L., Huang, G., Fan, J., Ma, X., Zhou, H., & Zeng, W. (2020). Hybrid extreme learning machine with meta-heuristic algorithms for monthly pan evaporation prediction. *Computers and Electronics in Agriculture*, *168*, 105115.

Wu, L., Zhou, H., Ma, X., Fan, J., & Zhang, F. (2019). Daily reference evapotranspiration prediction based on hybridized extreme learning machine model with bio-inspired optimization algorithms: Application in contrasting climates of China. *Journal of Hydrology*, *577*, 123960.