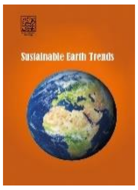
Sustainable Earth Trends

Journal homepage: http://sustainearth.sbu.ac.ir



Applicability of different ensemble techniques in enhancing water requirement simulations

Ahmad Jafarzadeh[a*], Abbas Khashei-Siuki[b], Ali Shahidi[c]

[a] Department of Water Engineering, Faculty of Engineering, University of Torbat Heydarieh, Torbat Heydarieh, Iran
[b] Department of Water Engineering, Faculty of Engineering, University of Torbat Heydarieh, Torbat Heydarieh, Iran & Department of Water Engineering, Faculty of Agriculture, University of Birjand, Birjand, Iran
[c] Department of Water Engineering, Faculty of Agriculture, University of Birjand, Birjand, Iran

ABSTRACT

Securing accurate insights into crop water requirements is crucial, particularly in arid regions and for key strategic crops. This study presents a novel applied methodology to evaluate the efficacy of ensemble modeling in determining the Saffron Water Requirement (SWR). The research aims to generate more reliable insights and enhance the accuracy of water requirement estimations for saffron, a vital crop in eastern Iran. The proposed plan includes establishing a rigorous testing process to evaluate the efficiency of various ensemble methods. Three significant ensemble classes— Ensemble Learning Machine (ELM), combination, and averaging techniques — were addressed to produce the new prediction of SWR. As such, a Decision Tree Regression (DTR) tool and six different experimental methods, serving as base models, were initially applied. The effectiveness of various ensemble methods was evaluated using statistical and qualitative tests. This plan included time series comparisons, key diagnostic indices such as RMSE and NSE, Absolute Error Decomposition (AED) analysis, and the Rate of Improvement (ROI). Results showed that applied ensemble systems are not only of high quality but also capable of presenting a skillful prediction of SWR compared to base models. Results revealed that the boosting procedure had a beneficial effect on DTR simulations, increasing them by more than 74 percent. Additionally, combining methods could enhance the base prediction by more than 75%.

ARTICLE INFO

Keywords:
Absolute Error Decomposition
AdaBoost
Kal-e Shur Sabzevar
Bias and Variance Errors
Collinearity
Inter-dependency

Article history:
Received: 09 June 2025
Accepted: 03 September 2025

*Corresponding author
E-mail address:
mnt.jafarzadeh@torbath.ac.ir
(A. Jafarzadeh)

Citation:
Jafarzadeh, A. et al., (2026). Applicability of different ensemble techniques in enhancing water requirement simulations, *Sustainable Earth Trends:* 6(3), (70-88).

DOI: 10.48308/set.2025.240361.1130

1. Introduction

Since water productivity is directly influenced by water consumption, obtaining a valid resolution of water requirements may be beneficial in protecting water consumption planning in the agricultural sector (Kouzegaran et al., 2020). This fact is crucial for strategic agricultural products, such as saffron (scientific name: *Crocus sativus L.*), in arid and semi-arid regions. Saffron farming is a key occupation in rural regions (Koocheki et al., 2016). Due to its high economic returns, it is often referred to as the "red gold crop" (Dessein et al., 2015; Leone et al., 2018). Additionally, this plant is recognized for its applications in industry, medicine, and food (Azgomi et al., 2021). The international position of Iran in saffron farming, the most contribution to production and cultivated area, is apparent (Shahnoushi et al., 2020; Ramezani et al., 2022) while saffron popularity is almost rooted in eastern regions, especially the Southern, Razavi, and Northern Khorasan provinces (Moshizi et al., 2023). Hence, the water consumption management of saffron in local development must be considered well. The accurate evaluation of Saffron Water Requirements (SWRs) is a crucial fact that motivated the researchers to undertake a diligent effort to enhance the quality of SWRs simulation. The literature highlights that Iranian scholar are pioneers in SWRs Studies (Alizadeh et al., 1997; Alizadeh, 2006; Azizi-Zohan et al., 2008; Sepaskhah & Yarami, 2009; Yarami et al., 2011). They started to evaluate the SWRs in Khorasan and Shiraz provinces by implementing local lysimetric laboratories. For example, Azizi-Zhohan et al. (2008) or Sepaskhah and Kamgar (2009) conducted studies to understand how irrigation methods and intervals influence productivity. Since then, numerous studies have investigated various aspects of saffron's water requirements (Maleki et al., 2011; Jafarzadeh et al., 2015; Shamsabadi et al., 2016; Fallahi and Mahmoodi, 2018; Koocheki et al., 2020). Determining the water requirements for regions without laboratory facilities is typically performed using experimental models and pre-evaluated crop coefficients. Many Evapotranspiration (ETO) models are widely used to estimate SWRs for these regions. However, employing pre-evaluated crop coefficients, suggested for standard conditions, along with simplifying assumptions of ETO models, propagates the uncertainty in water requirements simulation.

Despite all efforts and valuable advancements in estimating potential evapotranspiration and water requirements, it is essential to consider that a single model-based simulation should always be accompanied by a specified uncertainty component driven by the model’s inherent weaknesses (Huang et al., 2022). This has motivated a profound enthusiasm for the application of multi-models, coincidentally in different research areas in recent years. Currently, ensemble modeling is one of the most widely used and arguably the most recent advancements considered practically in studies aimed at improving results (Paul et al., 2023). Generally, ensemble modeling can be categorized into three groups: combination techniques, model averaging, and ensemble learning machines (ELM) (Zhou, 2019). The combined ensemble techniques include the methods that combine the outputs of different single models. The concept behind these methods is to overcome total uncertainty by considering the measurement average and the variance averaged across individual models. In this class, the standard methods that have been frequently cited are Simple Model Averaging (SMA) and Multiple Super Ensemble Modelling (MMSE). The second class, i.e., model averaging, focuses on reproducing a weighted average of different single models. A key process in these methods is determining the contribution of each input from competing models. Directly, the Weighted Averaging Modeling (WAM) and Bayesian Model Averaging (BMA) have received significant attention from researchers. The fundamental concept in the WAM method is to estimate contribution weights using the Least Squares Method, while BMA determines the weights based on probabilistic facts and the likelihood function. A literature review indicated the good performance of these methods in groundwater media (Jafarzadeh et al., 2021), rainfall-runoff water (Samadi et al., 2020), and infiltration contexts (Sang et al., 2023). However, how practical are these methods in enhancing SWR simulation?

The ELM class emphasizes the fact of boosting the outputs of a single learning machine. The main difference between ELM and the two mentioned ensemble methods is that ELM involves a single model, whereas the combination and model averaging techniques utilize multiple single models. Indeed, ELM attempts to generate an enhanced output through extensive resampling of the learning machine’s output. The Boosting, Bagging, and Voting algorithms are most commonly employed in ELM methods. The promising application of ELM has been reported in the reservoir permeability prediction (Otchere et al., 2022) surface water process (Li and Yang, 2023), groundwater context (e. g., Avand et al., 2020; Mosavi et al., 2021); urban water supply and distribution (Xu et al., 2022) while the applicability of this methods in SWRs has been dismissed so far. In one of the rare studies, Zarei et al. (2021) employed the Bagging algorithm to enhance the simulation outputs of water requirements for barley under different climate conditions in Iran. Nevertheless, can ELM methods upgrade the SWR simulation? Is this class the best option for implementing ensemble modeling, or are other ensemble methods, such as model averaging and combining techniques, more effective than the ELM approach? A clear response to these questions requires more research. It may be accepted that there are limited studies in which all ELM algorithms have been examined concurrently for their impact on the water requirement of saffron.

In conclusion, the review of the performed studies confirms a severe research gap in the applicability of ensemble modeling that still needs to be addressed.

First, to the best of the authors' knowledge, there has been no open discussion about the relevance of different existing categories of ensemble modelling simultaneously to determine their scope, application, weaknesses, and strengths. More recent studies have focused on ELM methods, overlooking the concurrent comparison of available ensemble methods, such as model averaging and combining. This study addresses the aforementioned gaps and provides a guideline for selecting the most effective ensemble methods for future global studies. Indeed, the best possible candidate is introduced under different conditions.

Second, water requirement simulation has been largely dismissed in most studies that employ ensemble approaches, resulting in fewer studies being conducted entirely. In this study, the feasible improvement of water requirement predictions is achieved simultaneously through three classes of learning-ensemble tools, combination techniques, and model averaging, as detailed below.

Third, saffron is one of the most crucial export products in arid regions worldwide, especially in Iran, and it can be accounted for as one of the research topics of the current study. Due to the high cultivation area of this crop, studies on water requirements can play a crucial role in water planning. Ultimately, a practical guideline was established in this study to provide a straightforward approach for selecting the best ensemble techniques under various conditions. Furthermore, a Python open-source framework was provided, which can serve as a guideline for future scopes.

2. Material and methods

2.1. Study area description

Birjand, the center of Southern Khorasan in eastern Iran, has a cold and arid climate pattern characterized by low annual rainfall (<120 mm) and a high annual evaporation rate (2600 mm). Groundwater is considered the primary source of water supply demands, and most exploitation occurs in the agricultural sector.

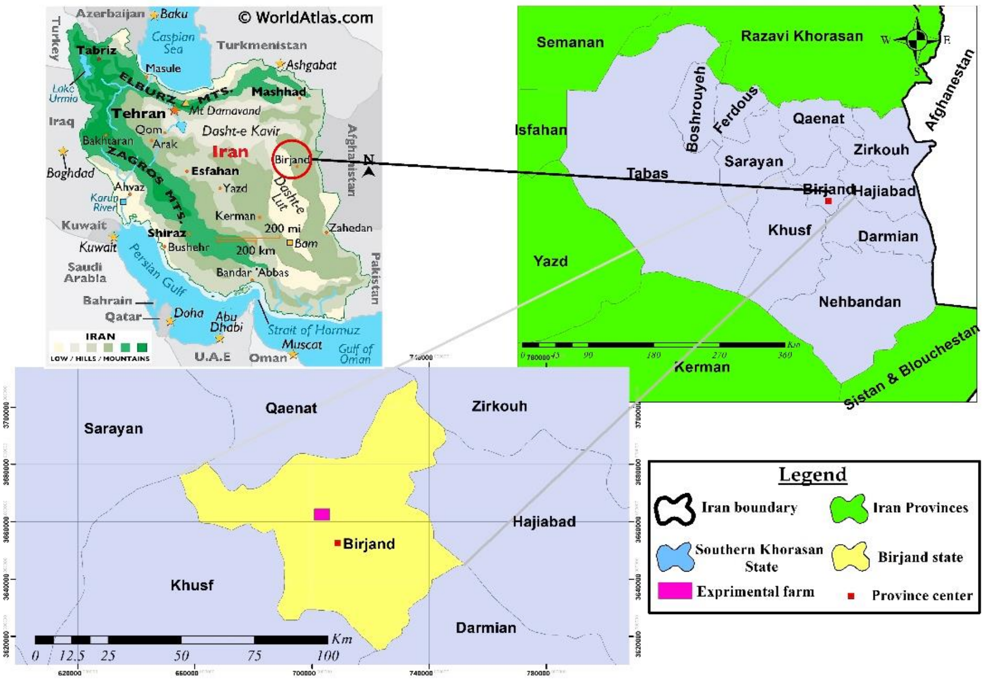


Fig. 1. A cartography view exhibiting the Birjand state and lysimeteric laboratory.

The current study utilized pre-recorded lysimeter data to estimate SWRs and their crop coefficients. A lysimeteric laboratory located in the Agriculture Faculty of the University of Birjand ($32^{o}\,53^{'}\,N, 59^{o}\,7^{'}\,E, 1480\,msl$) was considered to record daily moisture context on six lysimeters (see Fig. 1). Based on water metering, the total volume consumed during the growing period was 280 mm, considered for each lysimeter. The SWRs' data for the second year can be calculated given the following (Eq. 1):

$$ET_C = I + P - D \pm \Delta S \qquad (1)$$

Where I, P, D, and ET_C represent the irrigation volume (mm), rainfall volume (mm), drained volume (mm), and actual water requirement (mm). Also, moisture content was indicated through ΔS (mm). Khashei-Siuki et al. (2020) calculated the 10 days-SWRs for each lysimeter. Fig. 2 demonstrates the crop coefficients of saffron for the second year.

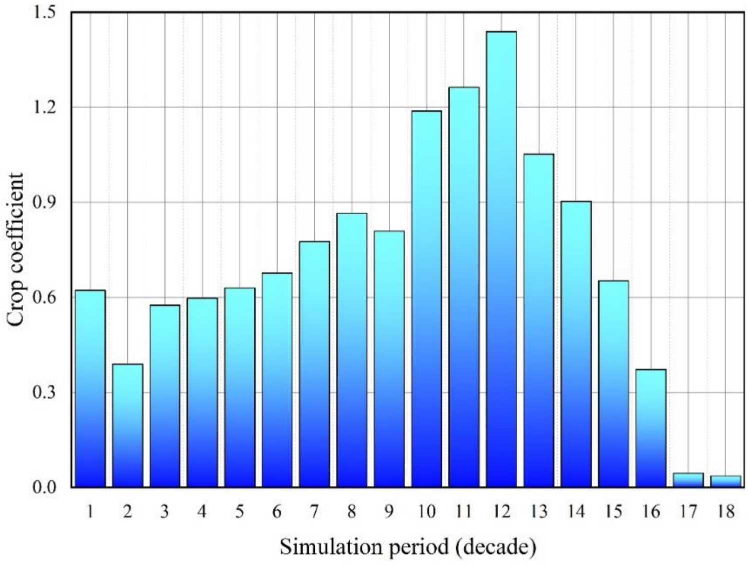


Fig. 2. Bar chart of evaluated saffron crop coefficient in the second year in Birjand experimental farm (Adopted from Khashei-Siuki et al., 2020).

The current study analyzed the interior concept of the investigation site through physical and chemical metrics presented in Tables 1 and 2. Furthermore, the required considerations regarding pre-processing practices, including handling missing values, eliminating outliers, and verifying stationarity assumptions, were addressed and confirmed.

Table 1. Describing the results of taking samples concerning the structure and chemical contents in the Birjand investigation site.

Depth	Texture	Bulk density	θ_{FC}	θ_{pwp}	Ec	pH	Organic carbon
cm	-	$gr.cm^{-3}$	%	%	$ds.m^{-1}$	-	
0-30	Loam	1.4	17.9	11.8	8.12	7.8	0.49

Table 2. Chemical components of the used water in the Birjand investigation site.

Component	Ec	pH	SAR	Ca^{++}	Mg^{++}	Na^{+}	K^{+}	Cl^{-}	$Hcom^{-}_{3}$	Com^{2-}_{3}	SO^{2-}_{4}
Unit	$ds.m^{-1}$	-	%	$meq.lit^{-1}$	$meq.lit^{-1}$	$meq.lit^{-1}$	$meq.lit^{-1}$	$meq.lit^{-1}$	$meq.lit^{-1}$	$meq.lit^{-1}$	$meq.lit^{-1}$
Magnitude	1.4	8	7.5	3.5	2.7	13.3	0.08	8.8	4.9	0	6.4

2.2. Ensemble modeling

Ensemble modelling involves mathematical procedures where the outputs of one or more base input models are combined to generate an enhanced outcome. The primary incentive for ensemble modelling execution is to reduce the total uncertainty of the prediction (Zhou, 2019). This study employed three ensemble methods, including an ensemble learning machine (Boosting and Bagging), combination techniques (M3SE), and model averaging (WAM). Additionally, a detailed description of the ensemble methods used is presented here.

2.2.1. Ensemble Learning Machine (ELM)

ELM is a form of a hybrid learning machine primarily introduced for classification targets (Nilsson, 1965), with the significant engine being repetitive sampling in this context. The ELM principle is to train multiple learners and aggregate their outputs into a work with better accuracy. The description of significant groups of ELM used in this study has been presented in the following:

Boosting: A boosting algorithm refers to methods that enhance weak learning algorithms to strengthen them. The idea of boosting is to correct the mistakes the weak learner makes. In summary, a repetitive adjustment is applied to several sequential learning machines to recognize error prediction, and the final prediction is obtained by combining the outputs of these learning machines. Freund and Schapire introduced the AdaBoost (Adaptive Boosting) algorithm in 1997. The Gradient Boosting and eXtreme Gradient Boost (XGBoost) algorithms also belong to the Boosting category, which is not discussed in this context. This algorithm begins by employing a learning machine (the first learner) to predict the original dataset, assigning equal weights to every record. In the second round, the algorithm adjusts the previous training outcomes, assigns more weight to uncorrected predictions, and subsequently applies the second learner. It leads to the conclusion that the second learner will most likely avoid making the mistake of the first one, despite its error. If this procedure is repeated, it is expected that the mistake made by the second learner will not occur in the third machine. This process will continue until a high level of training is achieved. Finally, AdaBoost combines the predictions of several learners to present an enhanced forecast (Fig. 3). The final prediction is obtained by a weighted average, where the weights are calculated based on the learner's performance.

Bagging: In Bagging, multiple learners are assembled and trained simultaneously, while in Boosting, this process is accomplished sequentially. The photograph below exhibits the distinction between Bagging and Boosting. The Bagging procedure for ELM application can be described as follows: First, bootstrapping determines the pre-specified number of training sub-data through the bootstrap method, with replacement. Second, training involves instructing a pre-specified number of learners on the samples generated in the previous stage. Third, aggregating: this involves combining the learners' predictions into a single prediction. In the aggregating step, the final model is generated based on hard voting, also known as majority voting or soft voting (i.e., the average of the predictions). This study employed Decision Tree Regression (DTR) as the base learner to simulate SWR (see Section 2.2.4 for more explanation).

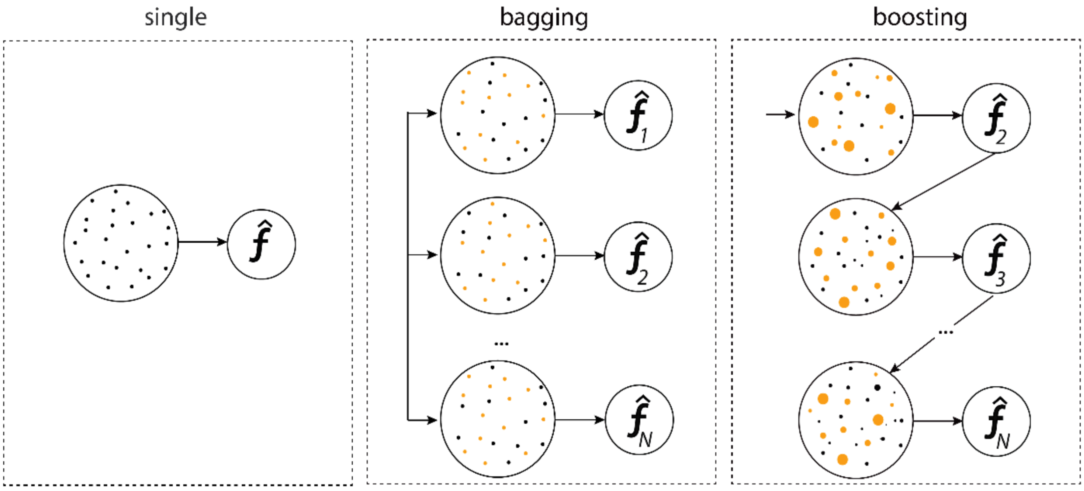


Fig. 3. A schematic view describing the Boosting VS Bagging context (Extracted from https://pluralsight2.imgix.net).

2.2.2. Combination techniques

These ensemble methods are based on individual models, and the basic idea is to pool the output of the participating models to achieve a more advanced outcome. This study employed the M3SE technique because it can be considered the last method presented in the combination techniques. Note that the participating items in this field were six different ETO models (described in the continuation) for predicting Etc.

M3SE: The leading development of M3SE is backed to Ajame et al. (2006), who added frequency mapping to an existing tool. The M3SE is given by as follows (Eq. 2):

$$WR^{t}_{M3SE} = \overline{WR}_{m} + \sum_{i=1}^{n} x_i.(WR^{t,i}_{Sim} - \overline{WR}^{i}_{Sim}) \qquad (2)$$

Where WR^{t}_{M3SE} is the M3SE projection, while x_i, $WR^{t,i}_{Sim}$, and $\overline{WR}^{i}_{Sim}$ denote, respectively, weight, the simulated water requirement, and the mean of the simulated water requirement. Moreover, $\overline{WR}_m$ reflects the average of the measured water requirement. More details are available from Ajami et al. (2006) and Jafarzadeh et al. (2021). Fig. 4 displays the applied steps of the M3SE methods.

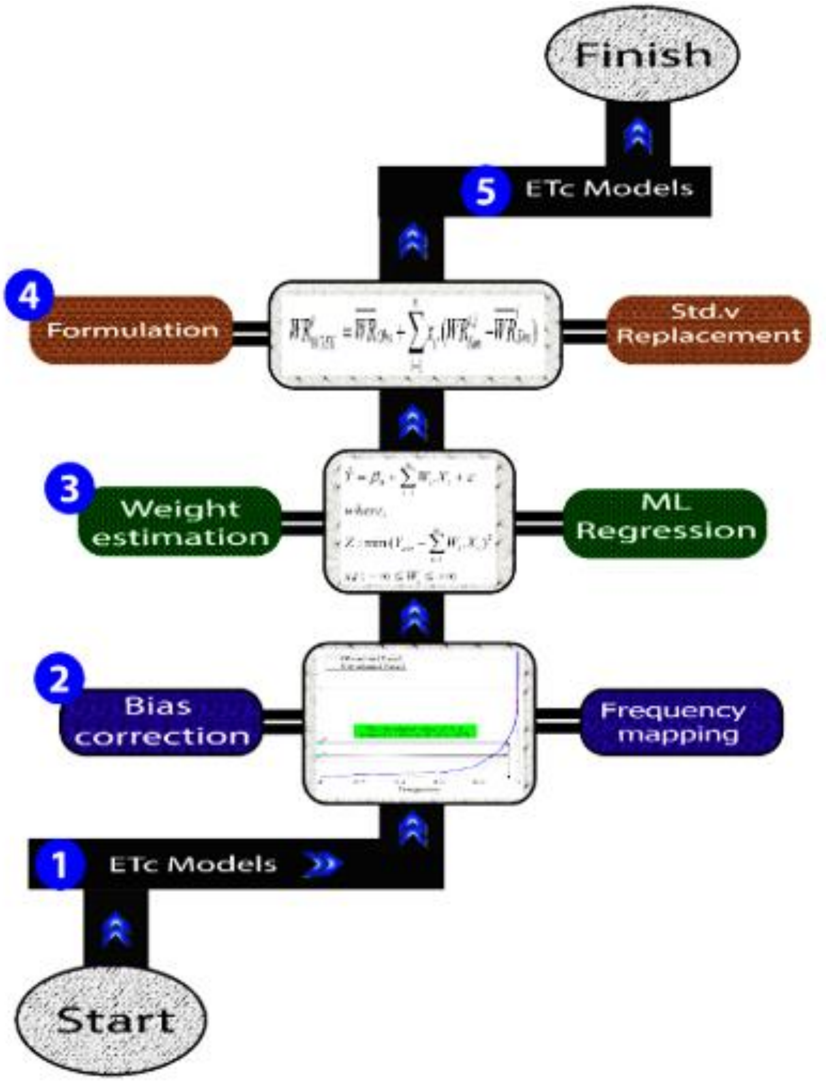


Fig. 4. A flowchart showing the applied steps of the M3SE combined method (Jafarzadeh et al., 2021).

2.2.3. Model averaging

WAM: This process produces a weighted average output. The same strategy can be adopted to estimate the quantity of weights, provided that they must be greater than zero and their sum must be unique.

2.2.4. Base input models

This research employed DTR as the base learner for an ensemble learning class, including boosting and bagging. Additionally, the six experimental models were considered as input models for combining and averaging classes (note that phrases such as 'participating', 'base', or 'input' models are used hereafter to refer to the DTR and ETO models). The required information about the methodology of the input models used in this study is available in the following section.

Decision Tree Regression (DTR): A DTR is a decision tree hired for time-series prediction (i.e., continuous simulations instead of discrete ones). The applied steps of DTR are described in well-known studies (e.g., Xu et al., 2005; Tso and Yau, 2007). Whoever, for exciting reading, a concise description is presented as follows: 1- The splitting or branching of input features. The essential context is to find the point for each independent variable at which it can be split into two parts, so that the Mean Squared Error (MSE) is minimized at that point. The average for sorted input data is compared with the standard of the target variable, and MSE is then calculated for each part. This process is performed repetitively to determine the best point. 2- The Root Node (RN) was determined by measuring impurity strategies. In classification, several well-known methods, including Entropy, Information Gain, and the Gini index, were employed. In regression and continuous contexts, a Standard Deviation Reduction (SDR) is recommended to identify the essential input feature (Eq. 3).

$$\begin{aligned} &SDR^1 = SD_t - SD_{tx}^1 \\ &where, \\ &SD_{tx}^1 = SD_1 . \frac{N_1}{N_t} + SD_2 . \frac{N_2}{N_t} \end{aligned} \qquad (3)$$

Where SD_t is the standard deviation of the target, while SD_{tx}^1 denotes the split standard deviation of the first input feature? Also, SD_1 and SD_2 indicate the standard deviation for the first and second parts of the input feature variable, while N_1 and N_2 indicate the number of data points for each divided part. The priority of input features is determined through their SDR.

3. The structure of DTR is first built based on the RN and its split point, where data are divided into two branches, depending on whether they are greater than or less than the split point. This branching is then continued based on the second input variable. This process will continue until more branching is not possible. Additionally, specific criteria govern branching, including variance and the number of recalled data points. The last level of division is called a leaf node; nodes that include two or more sub-nodes are referred to as parent nodes, and their sub-nodes are called child nodes. The depth of DTR is also determined based on the branching levels. An elimination diagram is presented in Fig. 5. In this study, DTR was implemented using four climate variables (i.e., temperature and humidity at maximum and minimum levels) as input features and SWRs as the target.

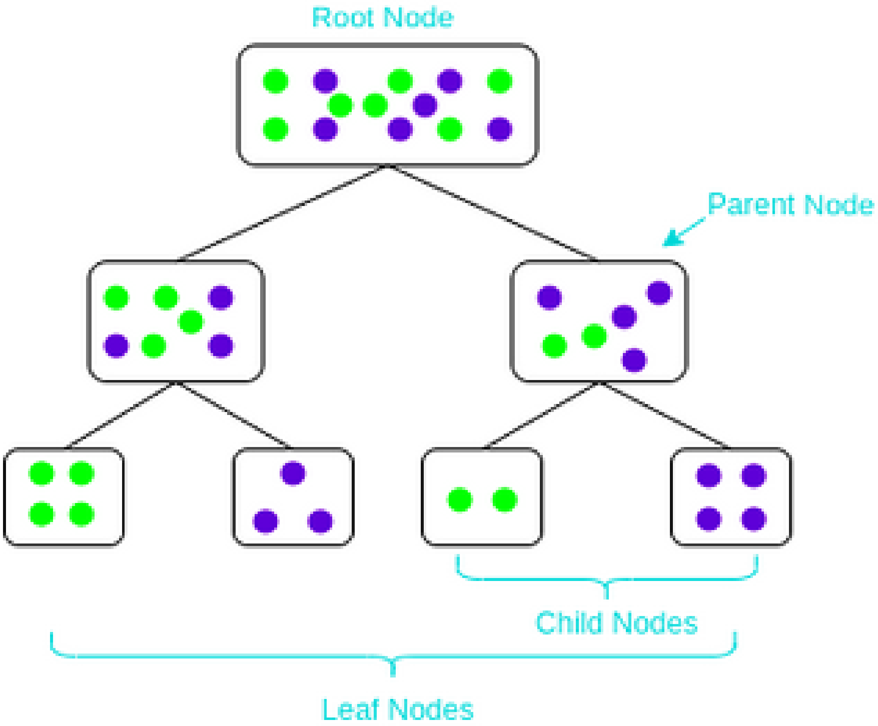


Fig. 5. A schematic view explaining the terminologies of standard DTR (Adopted from Tso and Yau, 2007).

Experimental evapotranspiration methods: This study employed six empirical models (i.e., FAO Penman-Monteith ‘FPM’, Priestly-Taylor Method ‘PTM’, Hargreaves Method ‘HM’, Turc Method ‘TUM’, Jensen-Haise Method ‘JHM’, Abtew Method ‘ABM’) to evaluate the 10-day ETc of saffron. The saffron crop coefficients were assessed based on reliable facts outlined in related works (Ghavamsaeidi Noghbi et al., 2020; Khashei-Siuki et al., 2020). Hence, the required explanation of these methods is presented as follows (Eqs. 4-9):

$$ET_{O\_FPM}=\frac{0.408\Delta(R_n-G)+\gamma\left[\frac{900}{T+273}\right]U_2\left(e_s-e_a\right)}{\Delta+\gamma\left(1+0.34U_2\right)} \quad (4)$$

$$ET_{O\_PTM}=\alpha\frac{\Delta(R_n-G)}{\Delta+\gamma} \quad (5)$$

$$ET_{O\_HM}=0.0022R_a*\sqrt{T_{\max}-T_{\min}}.(T+17.8) \quad (6)$$

$$ET_{O\_TUM}=0.013\left(\frac{T}{T+15}\right).(23.9R_a+50).\beta \quad (7)$$

$$ET_{O\_JHM}=\frac{(0.025T+0.08)R_a}{\lambda} \quad (8)$$

$$ET_{O\_ABM}=0.53\frac{R_a}{\lambda} \quad (9)$$

The definitions and descriptions of all parameters (Ra, Rn, T, Tmax, Tmin, Lambda, U, and delta) can be found in Allen et al. (1998). The different outputs of empirical ETO models will then be multiplied by saffron coefficients to give SWRs.

2.3. Model evaluation

This section explains how the various ensemble modeling techniques used in this study were examined. To define the proficiency level of each model, we designed comparison experiments that included time series comparisons, statistical diagnostic indexes (RMSE and NSE), and Absolute Error Decomposition (AED) analysis. Additionally, the skill of ensemble modeling techniques in improving SWR simulation was quantitatively calculated through the Rate of Improvement (ROI). This tool focuses on the extent to which the ensemble method has improved the outputs of individual models. This context can be given as follows (Eq. 10):

$$ROI=\frac{P_{EMT}-P_{SM}}{P_{SM}}\times 100 \quad (10)$$

Where, P_{EMT} and P_{SM} demonstrate the performance of ensemble techniques and single models, respectively. To calculate the ROI, the RMSE statistic was used to evaluate the performance of ensemble procedures and base methods.

2.4. Model setup

A general review of the implemented plan, conducted in separate consequential phases, is provided here. A meteorological dataset containing climate features (i.e., daily temperature, humidity, and wind speed) was first established during the simulation period, which spanned 180 days from November 8, 2020, to April 16, 2021. Additionally, the evaluated saffron crop coefficients and measured water requirement (ETc) were utilized to facilitate the ensemble model process. Then, the DTR machine and six different models of ETO were applied as input models. As such, the DTR machine was trained to compute the time series of SWR based on temperature and humidity (i.e., 10-day values of temperature and humidity were considered as input features, and SWR was assumed as the target). The first 130 data points (corresponding to 130 days from November 8, 2020, to March 17, 2021) were allocated for training the DTR model, and the remaining data (50 days from March 18, 2021, to April 16, 2021) were assigned for testing. Also, six different outputs of SWR were collected through various ETO models. Three different ensemble modeling approaches were used to duplicate a newly enhanced series of SWRs simulations. As such, the capacity of two enhancing techniques (boosting and bagging) was addressed in the ELM approach, and the effectiveness of the M3SE technique, in a combined manner, was demonstrated. Furthermore, the efficiency of model averaging was evident in the WAM scenario (Fig. 6).

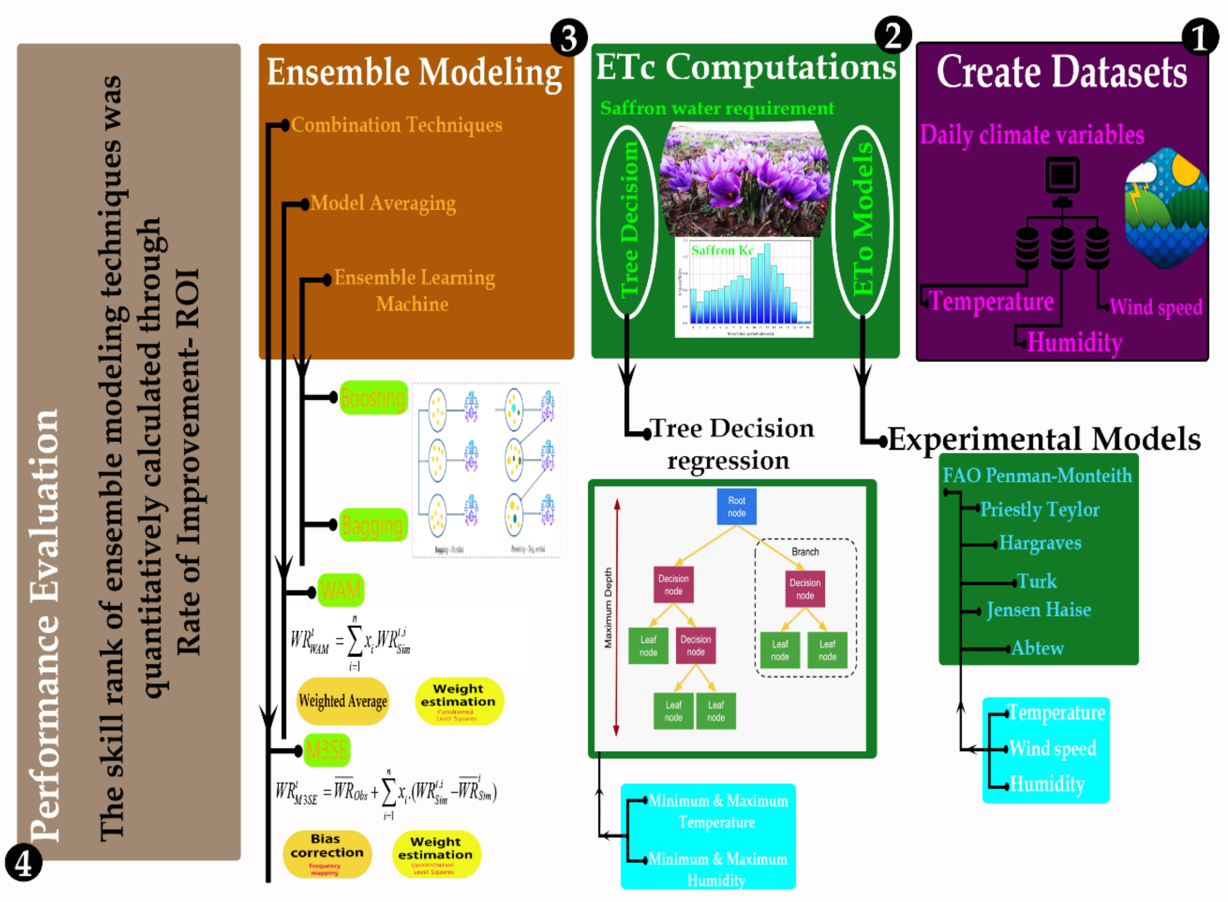


Fig. 6. The visual workflow explaining the leading steps of the current study.

3. Results and Discussion

3.1. Performance appraisal of base models

The ability of the two primary base models, DTR and ETO models, was first inspected.

3.1.1. Experimental evapotranspiration method

This section presents the order of methods in estimating SWRs based on the gained ability. Results from the survey (Table 3) confirmed the excellent prediction of the ABM method. In contrast, the estimated SWRs obtained by other methods do not exhibit an excellent match to the actual recorded data based on two criteria.

Table 3. Statistical results for Experimental models set.

Model	NSE	RMSE-mm	MAE mm
PTM	0.71	5.67	3.82
FPM	0.59	6.77	2.90
HM	0.69	5.96	3.28
TUM	0.53	7.30	3.86
ABM	0.96	2.16	0.94
JHM	0.66	6.18	3.78

Fig. 7 delivers a graphical comparison of the 10-day recorded and predicted water requirements for the four best models (for brevity and better display). The actual SWR recorded was 365 mm, and the derived results indicate that more ETO models have reached this amount of water requirement, with a relative consensus among all models. Nevertheless, a more precise survey clarifies more facts about their performance.

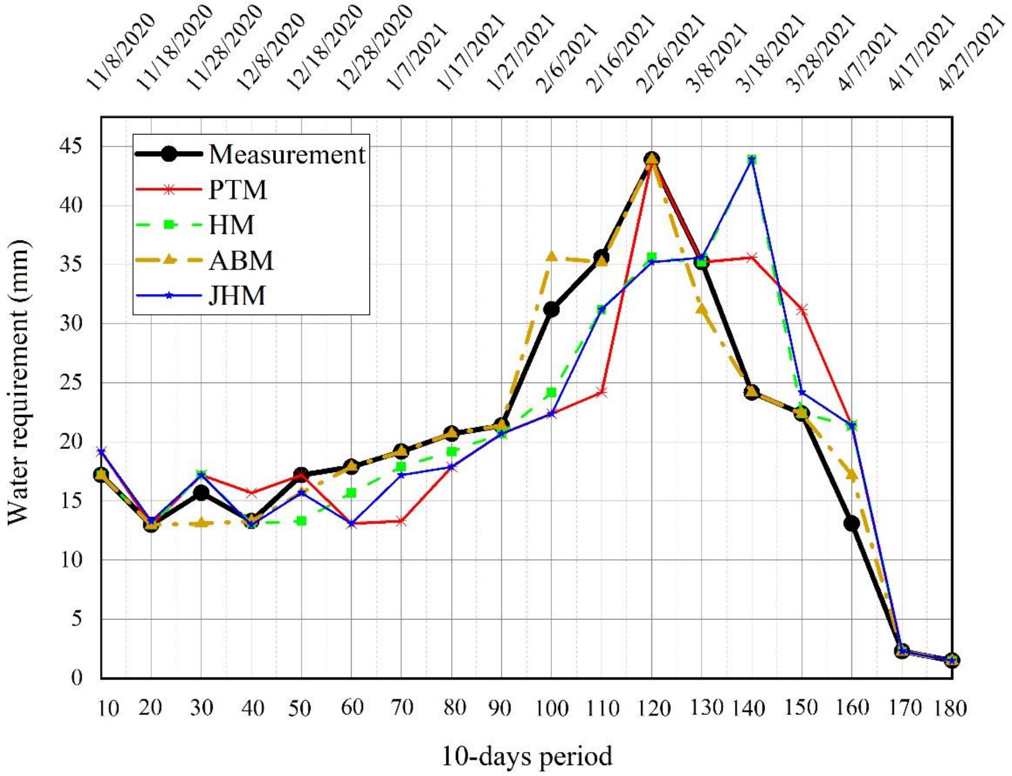


Fig. 7. Illustration of the time series of the experimental outputs versus the measured saffron water requirement.

Dividing the simulation period (180 days, from November 8, 2020, to April 16, 2021) into two parts of 90 days, it can be inferred that all ETO models exhibit better performance in the first part than in the second part. This matter can be Fig out by considering error decomposition for the two mentioned parts (see Table 4). This table represents the MAE and total absolute error that occurred in the first and second parts to discover a more accurate level of the ETO models' ability. The actual SWR in the first and second parts are 155.6 and 209 mm, respectively, while these values were estimated to be 151.5 and 213.5 mm by ABM, 147.3 and 217.7 mm by PTM. The results, as represented in Tables 3 and 4, confirm that the ABM model is the most proficient model, with PTM and HM placed in second and third places, respectively. Additionally, Turk's prediction is the worst.

Table 4. Results of AED (mm) for all experimental models, separated by rate of different parts.

Part	PTM	FPM	HM	TUM	ABM	JHM
1	20.1	8	11.3	18.1	4.1	15.9
2	48.7	44.2	47.7	51.3	12.9	52.1
Whole period	68.8	52.2	59	69.4	17	68

In general, the radiation and temperature-based models showed better performance for reflecting SWRs in the case studies. This fact aligns well with the consequences outlined by other researchers worldwide (e.g., Shi et al., 2022, in the Wet Tropics of Queensland, Australia; Salam et al., 2020, in Bengal areas, eastern regions of India; and Monteiro et al., 2021, in Brazilian areas). To illustrate, Salam et al. (2020) compared the proficiency of various empirical processes to simulate the ETO in Bangladesh. They reported that the empirical radiation model of ABM is successful in producing ETO in terms of accuracy and reliability, and its application was recommended for the same sub-regions. Monteiro et al. (2021) also noted that the ABM model had promoted performance among 29 evapotranspiration estimators in Brazilian areas.

3.1.2. Learning Machine- Decision Tree Regression (DTR)

The applicability of DTR in simulating SWRs is discussed here. This study considered the first 13 data points (130 days, from 8 November 2020 to 17 March 2021) for training the models, and the remaining data (50 days, from 18 March 2021 to 16 April 2021) accounted for the testing part. Also, some pre-processing practices, such as stationery and trend analysis, were examined and confirmed. Fig. 8 releases a time series of evaluated SWRs through DTR along with measured values. Additionally, the RMSE index was calculated for both the training and testing periods to assess DTR's ability to predict SWRs. As shown, evaluated SWRs have acceptable closeness to observed ones, indicating the good proficiency of DTR. The actual amount of SWR was 365 mm, while the TDR prediction gave 362 mm. The total RMSE and NSE were 3.06 mm and 0.92, respectively.

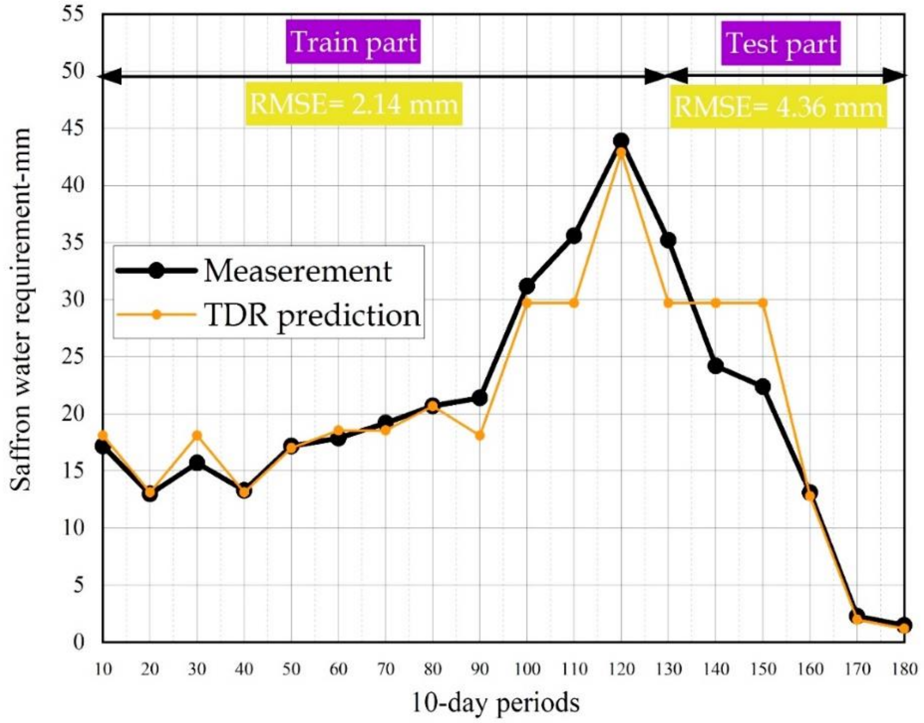


Fig. 8. A time-series plot comparing the measured SWRs versus DTR outputs.

A strong and consistent agreement exists between the DTR performance obtained in this work and studies conducted in related areas of water engineering (cf. Wei et al., 2019; Pekel et al., 2020). For example, Asadollah et al. (2021) stated that DTR outperforms SVR in predicting the quality context of the stream.

3.2. Performance appraisal of ensemble models

The results of learning-based ensemble methods indicated that the consequent training (i.e., boosting) improves the SWRs simulation compared to the iterative and parallel strategy (i.e., bagging). Fig. 9 shows the time series of recorded values against DTR concurrently, boosting, and bagging outputs. An over-prediction occurs in the first 60 days and the last 30 days of bagging production, while boosting predicted SWRs closer to the measured value, even more closely than DTR estimations. A higher accuracy of ensemble modeling can be better realized through diagnostic indices, where total RMSE and NSE were obtained as 0.78 mm and 0.99, respectively, for the boosting technique. From this perspective, it could effectively enhance the SWR simulation. The results for the bagging method were not confirmed due to its NSE of 0.88 and RMSE of 4.19 mm.

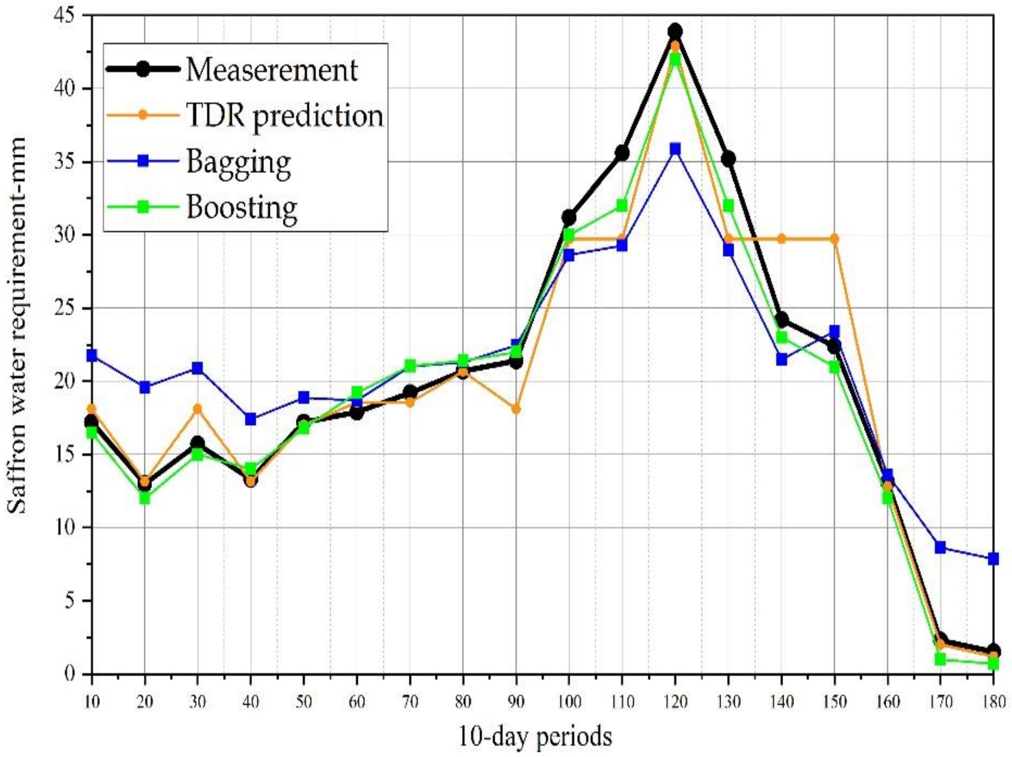


Fig. 9. Time series plot of SWR simulation indicating the effect of boosting and bagging ensemble algorithms in improving DTR outcomes.

The reason for the better results of the boosting algorithm lies in the nature of this method, where consecutive learning machines are organized to reduce the total bias. However, there are studies in which the proper performance of the bagging algorithm was reported (Chen et al., 2019; Salam and Islam, 2020). The potential answer is that bagging reduces variance, successfully eliminating overfitting that did not exist here. Generally, When the modeling performance in the test period is very different from the training and its accuracy is also not very favorable, the applicability of the bagging algorithm is likely more attractive.

For example, Salam and Islam (2020) attempted to test the skill of different ensemble methods, especially bagging, in simulating reference evapotranspiration through limited data. They concluded that the meta-based algorithm, bagging fashion, can reflect ETO in both training and testing splits. These results reveal that choosing bagging and boosting algorithms requires some pre-processing considerations affecting their performance. Thus, boosting is recommended when the learning machine goes with a bias error, and bagging is recommended when work has a variance error (Zounemat-Kermani et al., 2021).

Investigation is followed by addressing the performance evaluation of averaging and combining-based ensemble techniques. The results of these two methods indicate that the combining method operates better than the averaging-based ensemble in improving SWR simulation. The derived RMSE and NSE for M3SE method were 1.67 mm and 0.97, while those were reported 4.19 mm and 0.84 for WAM. This outcome points out that these two ensemble methods could upgrade the SWR predictions of ETO models strongly. Further examination is provided through time series test. Fig. 10 illustrates the original water requirement data compared to simulations improved by M3SE and WAM, and ABM (the best empirical methods). Results show that M3SE outperforms WAM, in improving the multi-model simulation because its output is slightly better than even ABM estimation.

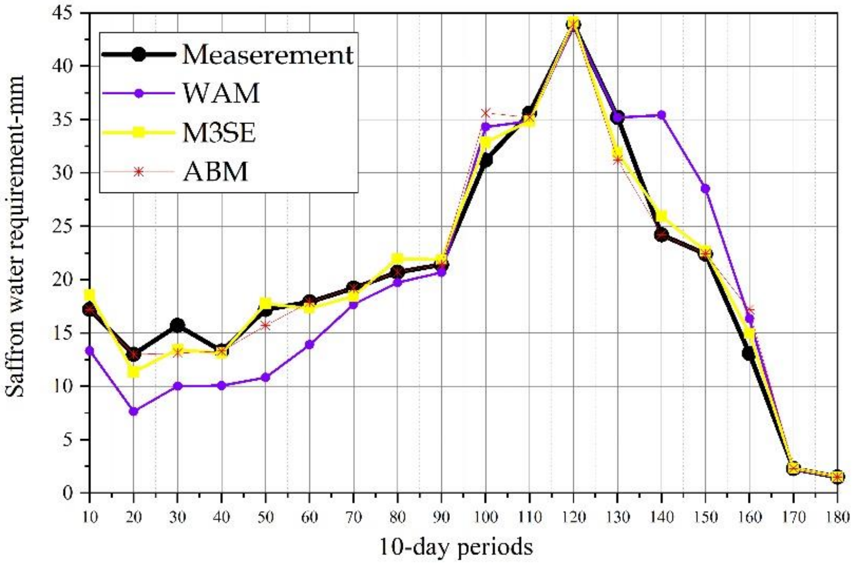


Fig. 10. Time series plot of SWR simulation indicating the effect of combining, M3SE, and averaging, WAM, ensemble methods in improving experimental outcomes.

This finding is related to the weighting strategy of WAM and M3SE. WAM allocated a set of positive weights whose summation must be one, while M3SE has no specified limitation in weighting. Further, the more sophisticated setting, such as frequency mapping, was implemented into M3SE, resulting in more accurate and skillful outputs than WAM. Fig. 11 shows the contribution of each participating method to the outcomes of WAM and M3SE in terms of allocated weights. As inferred, the WAM paid essential attention to ABM and PTM methods and eliminated the role of other ways. In contrast, M3SE has allowed all participants to contribute (except PTM). On the other hand, the estimated weights of M3SE have no relationship with the skill of each input model. For instance, M3SE considered a noticeable contribution for JHM, which had no significant performance reflecting the unreliable interpretation of prediction. This result is also very compliant with the studies that investigated the performance of combined methods in the fields of rainfall-runoff (Ajami et al., 2006), climate projection (Tebaldi and Knutti, 2007), and groundwater (Jafarzadeh et al., 2021). They found that the total uncertainty could be reduced under combining techniques, particularly the M3SE method. The better working was also expressed by Wang et al. (2009) for soil water monitoring.

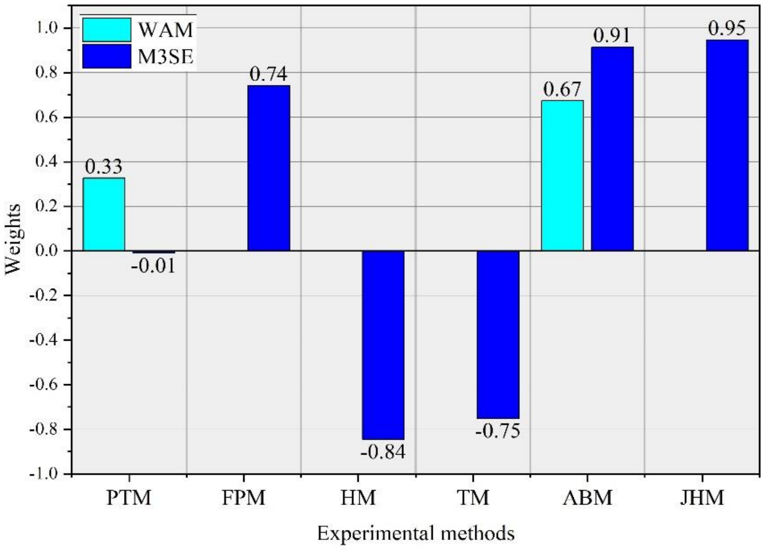


Fig. 11. A comparative graph showing the estimated weights of experimental methods for WAM and M3SE.

The above results explain the reason for less attention to M3SE compared to boosting and bagging methods, despite its good performance. A plausible cause for it can be an expansion of machine learning that motivates researchers to more using of these two algorithms. However, some inherent shortcomings, like the discussion in Fig. 11, prevent researchers from using this method. Many recent studies (e.g., Jafarzadeh et al., 2021), reported that despite the good ability of M3SE in bias cancellation, the explanation of input models' contribution is intricate. They stated that clarification of contribution is difficult in collinearity (i.e., a high correlation between input models). They offered to test a probabilistic tool, such as Bayesian Model Averaging (BMA). Therefore, it may be concluded that for combination inter-independence input models, applying M3SE is advised, and more alternatives, such as BMA, are recommended for collinearity situations.

3.3. Comparison of proposed methods

A good and practical judgement about applicability of different ensemble techniques can be drawn through simultaneously comparison. The reported accuracy criteria were collected in Table 5 to give a quick and precis check regarding different ensemble classes. Table 5 contains the NSE and RMSE for different categories, techniques, and input models. Concerning inserted values, it can be revealed that M3SE (combining ensemble method) upgraded strongly the quality of ETO estimations, such that it increased the NSE 0.28 (from 0.69 to 0.97) and decreased RMSE 4 mm (from 5.67 to 1.67). Also, boosting generated an improved SWR estimation such that NSE was raised 0.02 and RMSE was reduced 2.28 mm.

Table 5. Results of coincident comparison of different ensemble techniques in producing enhanced SWR simulations.

				NSE	RMSE (mm)
Ensemble classes	Combining and averaging	Techniques	WAM	0.84	4.19
			M3SE	0.97	1.67
		Input models	ETO models	0.69^{*}	5.67^{*}
	Ensemble learning	Techniques	Boosting	0.99	0.78
			Bagging	0.88	4.403
		Input models	DTR	0.92	3.06

* The average values were inserted.

Further discussion was proceed using error decomposition results. Fig. 12 supplies with a radar chart-based AED propagation for each ensemble method in which the amount of occurred error at any decade is recognized through radius quantity (i.e., major grid lines of 3, 6, 9, and 12 mm) and area of white polygon reflects the error propagation. For example, the occurred absolute error in 30 days for WAM and bagging are respectively, 5.8 and 4 mm almost. Based on this analysis, more AED multiplication occurred in WAM and bagging methods (more white area), while in M3SE and especially boosting, slight error prediction was taking place. To protect this finding, the RMSE and NSE values were inserted into Fig.12 to give a quick conclusion.

In proceeding, the quantity of the ROI index is presented in Fig. 13 as a doughnut pie chart defining ability of each ensemble technique in improvement of SWRs simulation. This graph gives the percentage of improvement and compares the strong of different ensemble methods simultaneously. Indeed, this analysis is another expression of derived results of Table 5, so it can be thought that this Fig. 13 is a concurrent comparison of different ensemble methods. The ROI results confirm that ensemble modelling can improve the quality of the final prediction irrespective of the type of ensembling. ROI analysis expresses that in combination and averaging techniques (with same input models), the execution of M3SE is more promoted. Because, the M3SE method improved the quality of ETO models' prediction more than WAM. For example, WAM enriched the FPM by more than 38 percent, while this progress in M3SE was near 75.33 percent. Further, the relevance of the boosted-ensemble learning machine may be almost similar to combined-ensemble methods because implementing boosting algorithm accompanied by DTR makes better predictions by more than 74 percent. These results strongly agree with related studies' findings (Ebrahimy et al., 2020; Sharafati et al., 2020; Mirzaei et al., 2021). For example, boosting algorithm's proficiency was reported by Kumari and Toshniwal (2021) for solar radiation forecasting.

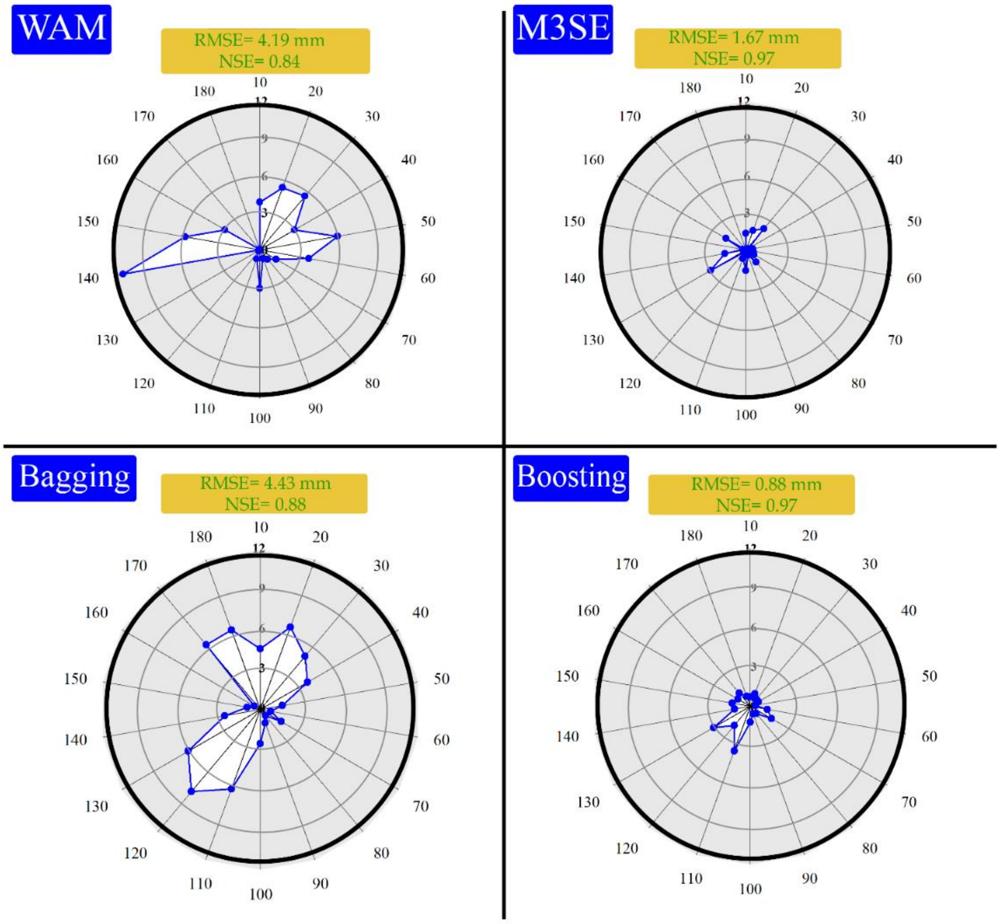


Fig. 12. A radar chart representing AED of SWR simulation during 10-days periods in separate of different ensemble methods: the radius shows the absolute error and white area represents the amount of uncertainty propagations.

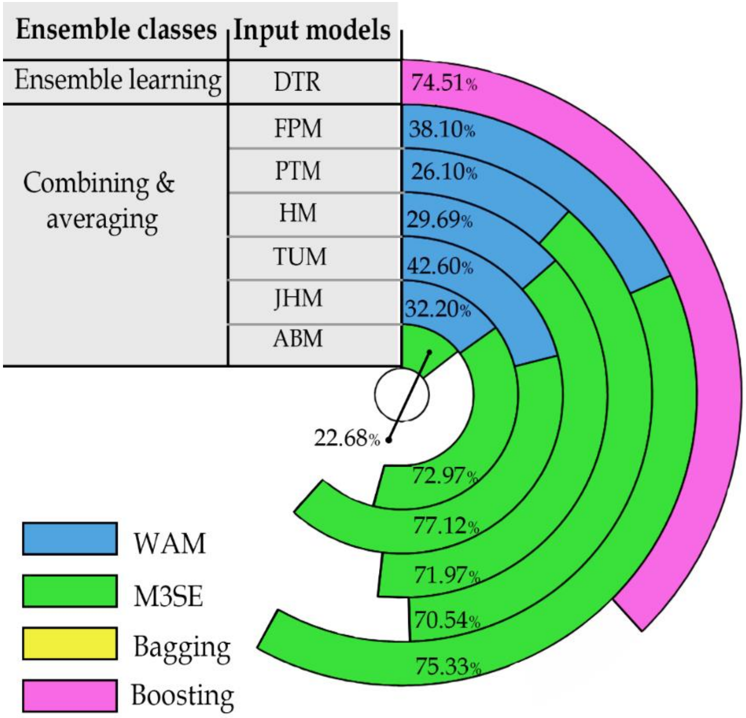


Fig. 13. A doughnut pie chart delineating ROI context of different ensemble techniques separately different classes and input models.

Overall, Boosting and M3SE ensemble methods have been recommended as the promising potentials for improving SWR simulation relying on derived results. Finally, a practical guideline regarding findings of current study is presented here to conduct proper direction for best selection of ensemble modeling. Determining the type of input models set, a relative path for selecting different techniques of ensemble modelling may be proposed through provided diagram in Fig. 14. Considering existing specifications in participating models and available algorithms (including frequency mapping, LSL, MCMC), the candidate techniques for ensemble modelling can be realized. This flowchart is unique and has been obtained relying on current study and may be upgraded in future, while it is surely a practical guideline for all simulation types.

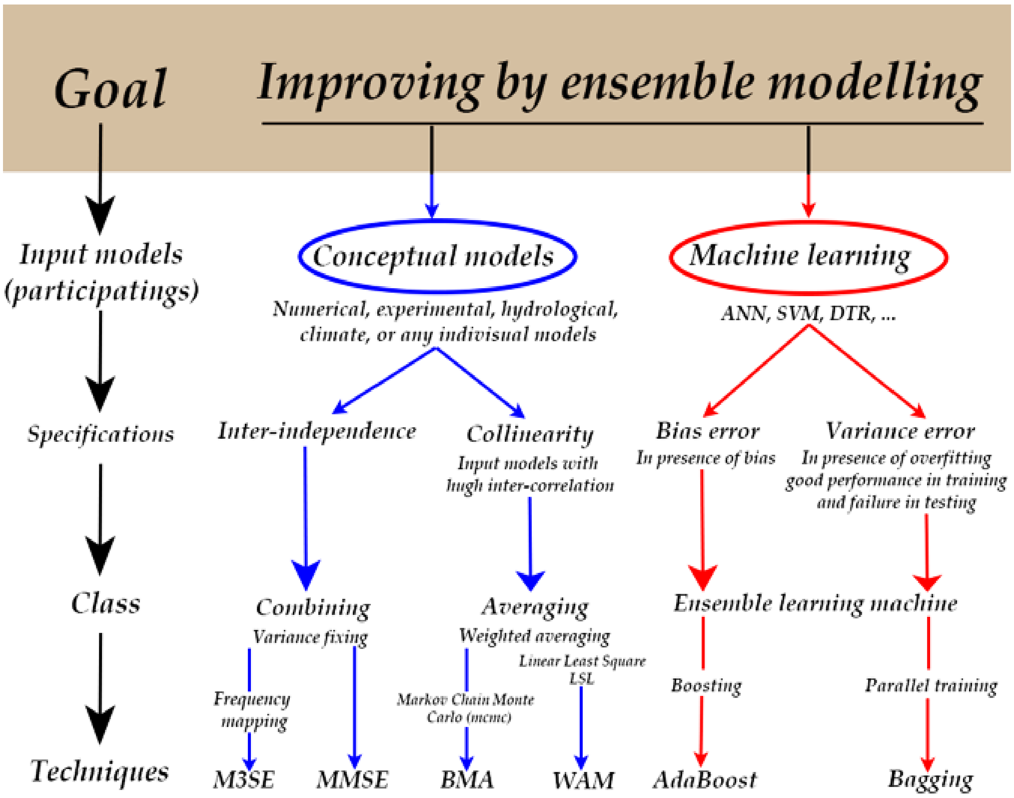


Fig. 14. A flowchart to select the most suitable techniques for ensemble modelling: Black arrows under Goal column (first left) show the direction of essential step while other blues and red arrows guide user from input model to best selection in last row (Techniques).

4. Conclusion

This paper explores the fact of affecting different approaches to dealing with ensemble modelling in the water requirement domain. A comparative survey was built to conceptualize the efficacy of ensemble modelling and clarify each technique's domain application, shortcomings, and strengths. To describe the proposed plan, the relevance of different ensemble methods was addressed to produce some developed series of SWRs in Birjand, south of IRAN. Different ETO models and a DTR machine were hired as input models to simulate the SWRs. Then, three primary ensemble techniques, named combination, averaging, and learning-based ensemble, were conducted to generate enhanced SWRs simulations. The performance evaluation of base models and ensemble techniques was addressed through time series comparison, AED analysis, statistical indices, and ROI. The evaluation results indicate that applied ensemble systems are not only of high quality but also capable of presenting a skilful prediction of SWR compared to base models. Overall, the findings of this study recommend the boosting and M3SE algorithms as attractive and good potential for enhancing the outcome of machine learning and experimental models. Because these two methods have been able to improve the accuracy of their base models by near 75%. The derived findings revealed that the boosted-based ensemble technique improved the DTR simulation by more than 74 percent. It is found that the use of bagging ensemble methods has a good outcome when the base learner is accompanied by variance error. Contrary to bagging, boosting accomplishes the ensemble process when the base learner faces a bias error. It is inferred that the combining ensemble method, M3SE, could better the performance of the ETO models between 22 and 75 percent for the best and worst participating models. However, using this method needs some concerns that must be examined further. Indeed, pre-processing must be committed to inspecting the existence of collinearity or inter-independence between input models. However, M3SE was strictly proposed when there is no inter-dependence between base models and BMA, more complicated method, was recommended in presence of collinearity. Some future research scopes are worthy of consideration to develop and promote the proposed plan. First, the current study examined the boosting algorithm based on the AdaBoost method. In contrast, there are some well-known algorithms (e.g., eXtreme and XGBoost) whose application was ignored here. Second, this study examined the limited ETO models as the base models, while abundant models there are for estimating referencing evapotranspiration. Further, we applied the DTR in an ensemble machine, and the examination of the rest learning machines is wholly felt. Third, since the current study focused only on limited data (second-year water requirement) and one specified crop (i.e., saffron), implementing the proposed plan for other crops with various growing patterns is precious to explore.

Acknowledgments

This research has been financially supported by the Saffron Institute, University of Torbat Heydarieh. Torbat Heydarieh, Iran. The grant number was 151230.

Author Contribution: 1- Ahmad Jafarzadeh: Conceptualization, Methodology, Investigation, Ensemble Modeling Implementation (i.e., Theory and Computations), Analysis of Results, Visualization, and Writing. 2- Abbas Khashei-Siuki: Conceptualization, Investigation Methodology, Investigation, Writing, Review, & Editing. 3- Shahidi, Ali: Conceptualization, Methodology, Investigation, Analysis of Results, Writing, Review, & Editing.

References

Ajami, N. K., Duan, Q., Gao, X. & Sorooshian, S., 2006. Multimodel combination techniques for analysis of hydrological simulations: Application to distributed model intercomparison project results. *Journal of Hydrometeorology*, 7(4), 755-768.

Alizadeh, A., 2006. Irrigation. In: Kafi, M., Koocheki, A., Rashed-Mohassel, M.H., Nassiri, M. (Eds.), Saffron, Production and Processing. Science Publishers,

Alizadeh, A., Mahdavi, M., Inanlou, M. & Bazari M.I. 1997. Potential evapotranspiration and crop coefficient of saffron. *Geographical Research Journal*. 12-46. 29-42.

Allen, R. G., Pereira, L. S., Raes, D. & Smith, M., 1998. Crop Evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. Fao, Rome, 300(9), D05109.

Asadollah, S.B.H.S., Sharafati, A., Motta, D. & Yaseen, Z.M., 2021. River water quality index prediction and uncertainty analysis: A comparative study of machine learning models. *Journal of Environmental Chemical Engineering*, 9(1), 104599.

Avand, M., Janizadeh, S., Tien Bui, D., Pham, V.H., Ngo, P.T.T. & Nhu, V.H., 2020. A tree-based intelligence ensemble approach for spatial prediction of potential groundwater. *International Journal of Digital Earth*, 13(12), 1408-1429.

Azgomi, R.N., Karimi, A., Zarshenas, M.M. & Jazani, A.M., 2021. The mechanisms of saffron (Crocus sativus') on the inflammatory pathways of diabetes mellitus: A systematic review. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 102365.

Azizi-Zohan, A., Kamgar-Haghighi, A.A. & Sepaskhah, A.R., 2008. Crop and pan coefficients for saffron in a semi-arid region of Iran. *Journal of Arid Environments*, 72(3), 270-278.

Chen, W., Hong, H., Li, S., Shahabi, H., Wang, Y., Wang, X. & Ahmad, B.B., 2019. Flood susceptibility modelling using novel hybrid approach of reduced-error pruning trees with bagging and random subspace ensembles. *Journal of Hydrology*, 575, 864-873.

Ebrahimy, H., Feizizadeh, B., Salmani, S. & Azadi, H. 2020. A comparative study of land subsidence susceptibility mapping of Tasuj plane, Iran, using boosted regression tree, random forest and classification and regression tree methods. *Environmental Earth Sciences*, 79, 1-12.

Fallahi, H.R. & Mahmoodi, S., 2018. Impact of water availability and fertilization management on saffron (*Crocus sativus L.*) biomass allocation. *Journal of Horticulture and Postharvest Research*, 1(2), 131-146.

Firoozjahantigh, M., Fakhri Alamdari, E. & Marzban, A., 2021. Investigating the effect of process-based instruction of writing on the IELTS writing task two performance of Iranian EFL learners: Focusing on hedging & boosting. *Cogent Education*, 8(1), 1881202.

Ghavamsaeidi Noghabi, S., Khashei-Siuki, A., Hammami, H., Shahidi, A. & Yaghoobzadeh, M., 2020. Determination of Evapotranspiration and Crop Coefficient of Saffron (*Crocus sativus L.*) by Lysimetric Method in the Dry-Desert Climate of Birjand. *Journal of Saffron Research*, 8(1), 161-172.

Huang, H., Band, S.S., Karami, H., Ehteram, M., Chau, K.W. & Zhang, Q., 2022. Solar radiation prediction using improved soft computing models for semi-arid, slightly-arid and humid climates. *Alexandria Engineering Journal*, 61(12), 10631-10657.

Jafarzadeh, A., Khashei-Siuki, A. & Shahidi, A., 2015. Modeling of climate change effects on saffron water requirement in south Khorasan province by GIS. *Journal of Saffron Research*, 3(2), 163-174.

Jafarzadeh, A., Pourreza-Bilondi, M., Khashei Siuki, A. & Ramezani Moghadam, J., 2021. Examination of various feature selection approaches for daily precipitation downscaling in different climates. *Water Resources Management*, 35(2), 407-427.

Khashei- Siuki, A. Shahidi, A., Behdani, M. A., Hjiabadi, F. & Shirzadi, F., 2020. Determination of Single and Dual Crop Coefficients of Saffron (*Crocus sativus L.*) in the First Year of cultivation. *Journal of Saffron Research*, 11(1), 108-123.

Koocheki, A., Fallahi, H.R. & Jami-Al-Ahmadi, M., 2020. Saffron water requirements. In Saffron (pp. 67-92). Woodhead Publishing.

Kouzegaran, S., Mousavi Baygi, M., Babaeian, I. & Khashei-Siuki, A., 2020. Modeling of the saffron yield in Central Khorasan region based on meteorological extreme events. *Theoretical and Applied Climatology*, 139(3), 1207-1217.

Kumari, P. & Toshniwal, D. 2021. Extreme gradient boosting and deep neural network-based ensemble learning approach to forecast hourly solar irradiance. *Journal of Cleaner Production*, *279*, 123285.

Leone, S., Recinella, L., Chiavaroli, A., Orlando, G., Ferrante, C., Leporini, L., ... & Menghini, L. 2018. Phytotherapic use of the *Crocus sativus L.*(Saffron) and its potential applications: A brief overview. *Phytotherapy research*, *32*(12), 2364-2375.

Li, S. & Yang, J., 2023. Improved river water-stage forecasts by ensemble learning. *Engineering With Computers*, 39(5), 3293-3311.

Maleki, M., Ebrahimzade, H., Gholami, M. & Niknam, V., 2011. The effect of drought stress and exogenous abscisic acid on growth, protein content and antioxidative enzyme activity in saffron (*Crocus sativus L.*). *African Journal of Biotechnology*, 10(45), 9068-9075.

Malik, A., Saggi, M.K., Rehman, S., Sajjad, H., Inyurt, S., Bhatia, A.S. ... & Yaseen, Z. M. 2022. Deep learning versus gradient boosting machine for pan evaporation prediction. *Engineering Applications of Computational Fluid Mechanics*, 16(1), 570-587.

Mirzaei, S., Vafakhah, M., Pradhan, B. & Alavi, S.J., 2021. Flood susceptibility assessment using extreme gradient boosting (EGB), Iran. *Earth Science Informatics*, 14, 51-67.

Monteiro, A.F.M., Martins, F.B., Torres, R.R., de Almeida, V.H.M., Abreu, M.C. & Mattos, E.V., 2021. Intercomparison and uncertainty assessment of methods for estimating evapotranspiration using a high-resolution gridded weather dataset over Brazil. *Theoretical and Applied Climatology*, 146(1-2), 583-597.

Mosavi, A., Hosseini, F.S., Choubin, B., Goodarzi, M., Dineva, A.A. & Sardooi, E.R., 2021. Ensemble boosting and bagging based machine learning models for groundwater potential prediction. *Water Resources Management*, 35(1), 23-37.

Moshizi, Z.G.N., Bazrafshan, O., Etedali, H.R., Esmaeilpour, Y. & Collins, B., 2023. Application of inclusive multiple model for the prediction of saffron water footprint. *Agricultural Water Management*, 277, 108125.

Nilsson, N.J., 1965. Learning machines McGraw-Hill. New York, 19652.

Otchere, D.A., Ganat, T.O.A., Gholami, R. & Lawal, M., 2021. A novel custom ensemble learning model for an improved reservoir permeability and water saturation prediction. *Journal of Natural Gas Science and Engineering*, 91, 103962.

Paul, S.N., Licata, R.J. & Mehta, P.M., 2023. Advanced ensemble modeling method for space object state prediction accounting for uncertainty in atmospheric density. *Advances in Space Research*, 71(6), 2535-2549.

Pekel, E., 2020. Estimation of soil moisture using decision tree regression. *Theoretical and Applied Climatology,* 139(3-4), 1111-1119.

Ramezani, M., Dourandish, A., Jamali Jaghdani, T. & Aminizadeh, M., 2022. The influence of dense planting system on the technical efficiency of saffron production and land use sustainability: Empirical evidence from Gonabad county, Iran. *Agriculture*, 12(1), 92.

Salam, R. & Islam, A.R.M.T., 2020. Potential of RT, Bagging and RS ensemble learning algorithms for reference evapotranspiration prediction using climatic data-limited humid region in Bangladesh. *Journal of Hydrology*, 590, 125241.

Salam, R., Islam, A.R.M.T., Pham, Q.B., Dehghani, M., Al-Ansari, N. & Linh, N.T.T., 2020. The optimal alternative for quantifying reference evapotranspiration in climatic sub-regions of Bangladesh. *Scientific Reports*, 10(1), 20171.

Samadi, S., Pourreza-Bilondi, M., Wilson, C.A.M.E. & Hitchcock, D.B., 2020. Bayesian model averaging with fixed and flexible priors: Theory, concepts, and calibration experiments for rainfall-runoff modeling. *Journal of Advances in Modeling Earth Systems*, 12(7), e2019MS001924.

Sang, J., Hou, B., Wang, H. & Ding, X., 2023. Prediction of water resources change trend in the Three Gorges Reservoir Area under future climate change. *Journal of Hydrology*, 617, 128881.

Sepaskhah, A.R. & Yarami, N., 2009. Interaction effects of irrigation regime and salinity on flower yield and growth of saffron. *The Journal of Horticultural Science and Biotechnology*, 84(2), 216-222.

Sepaskhah, A. & Kamgar Haghighi, A., 2009. Saffron irrigation regime. *International Journal of Plant Production*, 3(1), 1-16.

Shahnoushi, N., Abolhassani, L., Kavakebi, V., Reed, M. & Saghaian, S., 2020. Economic analysis of saffron production. In Saffron (pp. 337-356). Woodhead Publishing.

Shamsabadi, V., Far, A.M., Tohidi, R. & Mirzaei, S.M.J., 2016. Evaluation of water consumption productivity of saffron in Iran (case study: the province of Khorasan Razavi). *International Journal of Agriculture and Biosciences*, 5(3), 102-104.

Sharafati, A., Asadollah, S.B.H.S. & Neshat, A., 2020. A new artificial intelligence strategy for predicting the groundwater level over the Rafsanjan aquifer in Iran. *Journal of Hydrology,* 591, 125468.

Shi, L., Feng, P., Wang, B., Li Liu, D., Zhang, H., Liu, J. & Yu, Q.. 2022. Assessing future runoff changes with different potential evapotranspiration inputs based on multi-model ensemble of CMIP5 projections. *Journal of Hydrology*, 612, 128042.

Tebaldi, C. & Knutti, R., 2007. The use of the multi-model ensemble in probabilistic climate projections. *Philosophical transactions of the royal society A: mathematical, physical and engineering sciences*, 365(1857), 2053-2075.

Tso, G.K. & Yau, K.K., 2007. Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. *Energy*, 32(9), 1761-1768.

Wang, A., Bohn, T., Mahannama, S.P., Koster, D.R. & Lettenmaier, D.P., 2009 Multimodel Ensemble Reconstruction of Drought over the Continental United States. *Journal of Climate*, 22, 2694-2712.

Wei, L., Huang, C., Wang, Z., Wang, Z., Zhou, X. & Cao, L., 2019. Monitoring of urban black-odor water based on Nemerow index and gradient boosting decision tree regression using UAV-borne hyperspectral imagery. *Remote Sensing*, 11(20), 2402.

Xu, M., Watanachaturaporn, P., Varshney, P.K. & Arora, M.K., 2005. Decision tree regression for soft classification of remote sensing data. *Remote Sensing of Environment*, 97(3), 322-336.

Xu, Z., Lv, Z., Li, J. & Shi, A., 2022. A novel approach for predicting water demand with complex patterns based on ensemble learning. *Water Resources Management,* 36(11), 4293-4312.

Yarami, N., Kamgar-Haghighi, A.A., Sepaskhah, A.R. & Zand-Parsa, S. 2011. Determination of the potential evapotranspiration and crop coefficient for saffron using a water-balance lysimeter. *Archives of Agronomy and Soil Science*, 57(7), 727-740.

Zarei, A.R., Mahmoudi, M.R. & Shabani, A., 2021. Investigating of the climatic parameters effectiveness rate on barley water requirement using the random forest algorithm, Bayesian multiple linear regression and cross-correlation function. *Paddy and Water Environment*, 19(1), 137-148.

Zounemat-Kermani, M., Batelaan, O., Fadaee, M. & Hinkelmann, R., 2021. Ensemble machine learning paradigms in hydrology: A review. *Journal of Hydrology*, 598(March), 126266.