Journal of Hydrology 598 (2021) 126266

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Review papers

Ensemble machine learning paradigms in hydrology: A review

Mohammad Zounemat-Kermani[a,*], Okke Batelaan[b], Marzieh Fadaee[a], Reinhard Hinkelmann[c]

[a] *Department of Water Engineering, Shahid Bahonar University of Kerman, Kerman, Iran*
[b] *National Centre for Groundwater Research and Training, College of Science and Engineering, Flinders University, Adelaide, Australia*
[c] *Chair of Water Resources Management and Modeling of Hydrosystems, Technische Universität Berlin, Berlin, Germany*

A R T I C L E I N F O

This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Qiuhua Liange, Associate Editor

Keywords:
Committee machine
Random forest
Data mining
Soft computing
Hydroinformatics

A B S T R A C T

Recently, there has been a notable tendency towards employing ensemble learning methodologies in assorted areas of engineering, such as hydrology, for simulation and prediction purposes. The diversity of ensemble techniques available for implementation in hydrological sciences has led to the development and utilization of different strategies in the implementation. This review paper explores and refers to the advancement of ensemble methods, including the resampling ensemble methods (e.g., bagging, boosting, and dagging), model averaging, and stacking viz. generalized stacked, in different application fields of hydrology. The main hydrological topics in this review study cover subjects such as surface hydrology, river water quality, rainfall-runoff, debris flow, river icing, sediment transport, groundwater, flooding, and drought modeling and forecasting. The general findings of this survey demonstrate the absolute superiority of using ensemble strategies over the regular (individual) model learning in hydrology. In addition, the boosting techniques (e.g., boosting, AdaBoost, and extreme gradient boosting) have been more frequent and successfully implemented in hydrological problems than the bagging, stacking, and dagging approaches.

1. Introduction

Hydrological systems are complex and are characterized by processes and events whose dynamics depend on various direct (e.g., meteorological and environmental factors) and indirect (e.g., human interactions) interconnected phenomena. Hence, hydrological models are prone to conceptual and systematic errors and uncertainty in simulated results. Due to this fact, several different model structures, including physical, statistical, stochastic, mathematical, and numerical, have been developed and applied (Yuan et al., 2019). In the last two decades, the use of Machine Learning (ML) methods, which are nonlinear and Soft Computing (SC) tools for extracting characteristics, trends, or rules from datasets, have been rapidly increasing for data-intensive hydrological modeling problems (Govindaraju, 2000; Zounemat-Kermani and Scholz, 2013; Rajaee et al., 2020).

Where there is a desire to select a strategy for supervised learning of a system (e.g., hydrological event), two main categories of ordinary (standard) learning and ensemble learning can be distinguished. Exploring a literature review for the successful application of ML methods in hydrology asserts that ordinary learning techniques are used in most of the reported studies (Zounemat-Kermani, 2017; Alizamir et al., 2020a, 2020b). However, due to their higher efficiency in modeling, the applications of ensemble ML models in hydrological modeling have been considerably enhanced in recent years (Singh and

Abbreviations: AdaBoost, Adaptive Boosting; AICA, Averaging using Akaike's information criterion; ANFIS, Adaptive Neuro-Fuzzy Inference System; ANN, Artificial Neural Network; BGA, Bates-Granger model Averaging; BICA, Bayes' Information Criterion; BMA, Bayesian Model Averaging; CART, Classification And Regression Trees; CHAID, Chi-square Automatic Interaction Detector; DDM, Data-Driven Model; ELM, Extreme Learning Machine; EML, Ensemble Machine Learning; ENN, Ensemble Neural Network; ENR, Elastic Net Regression; ERT, Extremely Randomized Trees; ET, Extra-Trees; EWA, Equal Weights Averaging; GBRT, Gradient Boosted Regression Trees; GEP, Gene Expression Programming; GMDH, Group Method of Data Handling; GP, Genetic Programming; GRA, Granger-Ramanathan Averaging; LGP, Linear Genetic Programming; LSSVM, Least-Squares Support Vector Machine; MA, Model Averaging; MARS, Multivariate Adaptive Regression Spline; ML, Machine Learning; MLR, Multiple Linear Regression; MMA, Mallows Model Averaging; NF, Neuro-Fuzzy; PCA, Principal Component Analysis; RBFNN, Radial Basis Function Neural Network; RF, Random Forests; RSM, Random Subspace Method; RtF, Rotation Forest; SAM, Simple Arithmetic Mean; SC, Soft Computing; SGB, Stochastic Gradient Boosting; SVM, Support Vector Machine; SVR, Support Vector Regression; WA-ANN, Wavelet-Artificial neural network; WA-ELM, Wavelet- Extreme Learning Machine; XGB, eXtreme Gradient Boosting.

* Corresponding author.
E-mail address: zounemat@uk.ac.ir (M. Zounemat-Kermani).

https://doi.org/10.1016/j.jhydrol.2021.126266
Received 11 December 2020; Received in revised form 13 March 2021; Accepted 29 March 2021
Available online 1 April 2021
0022-1694/

Panda, 2015; Shin et al., 2017; Sun et al., 2020). Ensemble learning consists of various approaches based on different methodologies behind their nature. Instances include stacking methods (also known as committee machine approach), averaging methods (which can be considered as simple stacking ensemble methods without having a *meta*-learner), bagging, and boosting approaches (Dietterich, 2002; Zhou, 2009; Sun and Trevor, 2018). Detailed descriptions of commonly used ensemble methods in hydrology are given in Section 3.

It is stated and proven that ensemble learning can overcome related shortcomings of ordinary learning of ML models, including statistical problems (when the search space is too large for the available training data), computational problems (where the learning algorithm fails to find the optimum solution), and representation problems (when the learning algorithms lack appropriate fitness functions) (Dietterich, 2002; Abba et al., 2019; Zounemat-Kermani et al., 2020b). Indeed, several studies have already claimed the superiority of Ensemble Machine Learning (EML) to ordinary learning in diverse hydrological modeling problems (Yu et al., 2020). Nonetheless, there is no thorough and comprehensive study that investigates the state-of-the-art and scrutinizes the recent trend of ensemble learning in different areas of hydrology. On that account, the objectives of the study are, therefore, twofold: 1) to provide a survey on the applications of different EML techniques in various fields of hydrology, including surface hydrology (river and streamflow modeling, river water quality, debris flow, river icing, suspended sediment transport, and rainfall-runoff process), hydrogeology (groundwater modeling), and hydrological events (flood and drought modeling), and 2) to evaluate and critique the implementation of different types of ensemble models (e.g., stacking, bagging, and boosting) in hydrological modelling.

2. Rationale and contribution

In spite of the increasing employment of EML methods in hydrological sciences, to the authors' best knowledge, a comprehensive survey of the applications of ensemble learning models in the pertinent topics is still missing. Hence, the current review examines and summarizes scientific articles published over the last two decades, presenting the application of different types of EML, such as model averaging, stacking, bagging, boosting, and dagging. On the whole, we have studied and reported more than 160 peer-reviewed papers focusing on hydrological areas such as surface hydrology, hydrogeology, and extreme hydrological events in hydrological sciences. We expect this study to contribute to recommendations and problem-specific guidelines for future usage of EML models in dealing with hydrological events and phenomena.

3. Methodology

Ensemble learning is a type of machine learning paradigm that consists of a number of learners called base learners (also known as weak learners) who are trained and combined to analyze and address a variety of real-world problems. The process of learning from base learners is called "meta-learning"; therefore, EML methods are also known as *meta*-learning methods (Zhou, 2009; Zhang and Ma, 2012; Sharghi et al., 2018; Zounemat-Kermani et al., 2020a, 2020b).

The EML process generally consists of generating and combining base learners (Fig. 1). In a first step, the ensemble generation and base learners are created. Next, the ensemble pruning step removes some of the created functions from the first step. In the last step, the ensemble combination combines the chosen functions.

Generally, EML methods form a broad category of algorithms based on both heuristics and sound learning-theoretic principles (Kim et al., 2019). In the following subsections, the most prominent and important EML, which are developed and applied widely and successfully in water science and hydrological issues are briefly described.

3.1. Model Averaging

Model Averaging (MA) approach has gained a lot of attention as a simple and efficient EML in modeling hydrological processes. These techniques consist of different methods such as Simple Arithmetic Mean (SAM), Bayesian Model Averaging (BMA), Equal Weights Averaging (EWA), Bates-Granger model Averaging (BGA), Averaging using Akaike's information criterion (AICA), Bayes' Information Criterion (BICA), Mallows Model Averaging (MMA), and Granger-Ramanathan Averaging (GRA) (Arsenault et al., 2015).

BMA is one of the most efficient and well-known EML algorithms that is able to incorporate model uncertainty. In this method, the production of predictions by weighted averaging is based on sets of different competing models. In BMA, the weighted average of all the models is taken with the weights provided based on posterior probabilities. In the following, the process of the BMA algorithm is described in brief.

In the MBA method, if $M = \{M_1, \dots, M_k\}$ is one possible set of k models, the posterior probability (p) for the model M_k given the dataset D is defined as:

$$p(M_k|D) = \frac{p(D|M_k)p(M_k)}{\sum_{l=1}^{K} p(D|M_l)p(M_l)}, \quad (1)$$

where

$$p(D|M_k) = \int p(D|\theta_k, M_k)p(\theta_k|M_k)d\theta_k \quad (2)$$

where D denotes the observed dataset, $p(D|M_k)$ is the likelihood of model prediction M_k, $p(\theta_k|M_k)$ is the prior density of the vector parameter, and $p(D|\theta_k, M_k)$ is the likelihood (Duan et al., 2007; Wright, 2008).

3.2. Stacking

Stacking is a specific type of EML algorithm that can produce a strong learner from poorer ones (Wolpert, 1992). In order to achieve better performance, this integrated algorithm, uses a higher-level model to integrate lower-level models, and finally increases the predictive power of the classifier. Besides, this approach aims to minimize the errors of generalization by reducing bias and variance.

The stacking method processing has two levels or stages containing level-0 and level-1. Various base models (model-1 to model-K) are trained based on the initial training dataset in level-0, and the prediction of the response variable for the individual models is conducted in this

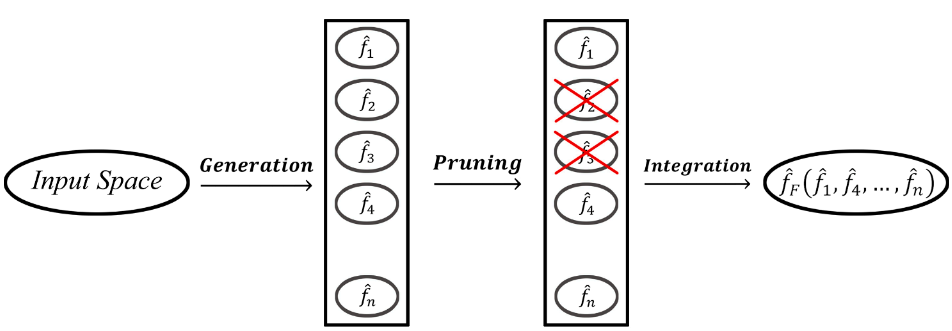

Fig. 1. The general procedure of setting up an ensemble learning strategy.

stage. Then the outputs of level-0 are combined from multiple base models into a single score (*meta*-model) and create the output of the next stage (level-1) to train an ensemble function (Sun and Li, 2020). Fig. 2 illustrates the standard two-layer stacking modelling process.

3.3. Bagging

The bagging algorithm (also known as bootstrap aggregating) was introduced by Breiman (1994). It is one of the simplest and earliest methods of EML techniques and is most suitable for problems with small training datasets. In this method, the bootstrap sampling method is used to acquire random subsets of data for training a set of original models with replacement. The majority vote is used for combining the individual output models obtained from bootstrap samples. The illustration of the Bagging algorithm process is shown in Fig. 3.

Bagging can be employed for improving the precision of machine learning approaches and used in regression and classification. This approach assists in the deduction of variance and enhancement of the robustness of the models by using decision trees; therefore, it is well-suited for original weak models with high variance (Polikar, 2012; Wu et al., 2020). There are several popular creative versions of bagging such as Random Forests (RF), Random Subspace Method (RSM), Extremely Randomized Trees (ERT), and Rotation Forest (RtF), which are explained in the subsequent subsections.

3.3.1. Random Forests

Random Forests (RF) are one of the best-known bagging algorithms that is suggested by Breiman (2001). Rfs are tree-based EML algorithms that use decision trees as base-learners, so that each tree is dependent on a set of random parameters. Each tree is created from a bootstrap sample of the primary dataset. RFs can be employed for either a categorical dependent parameter (response variable) for classification purposes or a continuous response variable, for developing regressive models. RFs, similar to other EML algorithms, combine multiple individual models to make predictions. Therefore, this method can be a better choice than a single decision tree since it decreases the over-fitting by averaging the result. The RF algorithm process has several stages: Firstly, bootstrap samples are randomly selected from the given original dataset. Then a decision tree is created for each sample, and the prediction result of each decision tree is obtained in this stage. Next, for each predicted result, the voting step is carried out. Eventually, the most popular prediction model is chosen as the final prediction result (Breiman, 2001; Cutler et al., 2012; Heddam, 2021a, 2021b).

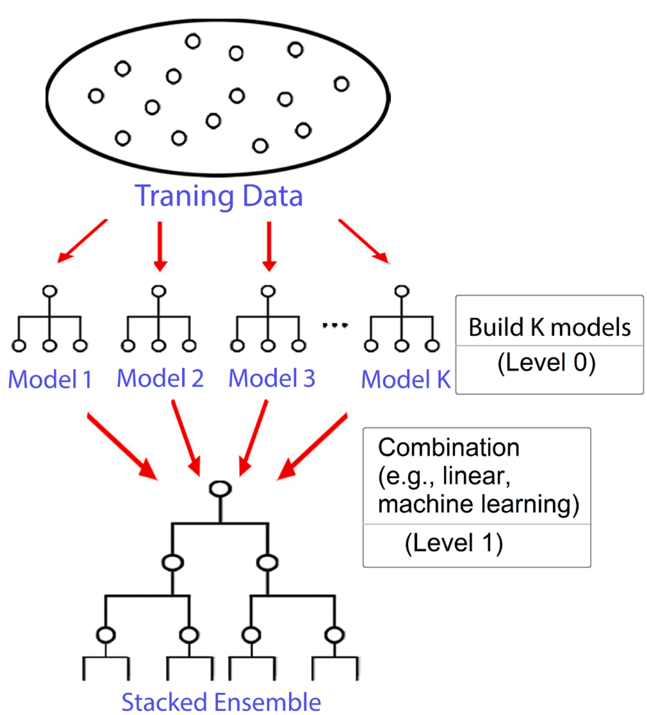

Fig. 2. Schematic illustration of the stacking ensemble learning methodology.

3.3.2. Random Subspace

The Random Subspace Method (RSM) is an ensemble classifier method suggested by Ho (1998). RSM is a parallel learning algorithm in which the process of each subspace creation is independent. For creating ensembles of learners in this method, modified feature space is used, and features in this feature space are sampled randomly for all base-learners with replacement. Next, whole samples are projected to a fixed subspace, and then these projected training samples are used to train a classifier. Finally, the outputs of the individual models are combined by the weighted majority voting factor in the model combination stage to lessen the error of RSM (Li et al., 2013; Mert et al., 2016; Salam and Islam, 2020).

3.3.3. Extremely Randomized Trees

Extremely Randomized Trees (also known as Extra-Trees (ET)) is a tree-based EML method (Geurts et al., 2006). This method performs similarly to the RF method and can be utilized for both classification and regression. In contrast with other tree-based ensemble methods (e.g., the RF method), in the ET method, the bootstrap sampling method is not applied and all trees are trained by the complete training dataset. Moreover, rather than selecting the best cut-point for each feature based on the local sample, the splits of attributes and cut-points are selected totally randomly for each of the features in accordance with the variable index and variable splitting (Geurts et al., 2006; Heddam et al., 2020).

3.3.4. Rotation Forest

Rotation Forest (RtF) is an EML classifier generation method with as its main purpose to create diverse and accurate classifiers inside the ensemble (Rodriguez et al. 2006). This method, similar to the concepts of bagging methods, uses the bootstrap sampling and trains decision trees independently based on creating a classifier ensemble using a feature extraction technique (such as Principal Component Analysis (PCA)). In Rotation Forest, in order to preserve diversity, extracting features is used for each of the base classifiers. To maximize individual accuracy, each of the base classifiers is trained on the entire dataset in the rotated feature space. The feature sets are split into subsets at random, then feature extraction (i.e., PCA) is employed to each of the subsets. Eventually, the final result is accomplished by combining the average outputs of all trees (Rodriguez et al., 2006).

3.4. Boosting

Boosting - a technique to enhance the accuracy and performance of machine learning algorithms - is a well-known ensemble learning. The basic concept of the boosting technique is adding new models to the ensemble sequentially. In this ensemble, weak learners (base learners) are effectively boosted to strong ones. Boosting is applied to curb the overfitting of decision trees. As a result, it helps to decrease variance and bias in an EML and increase the accuracy of predictions.

Similar to Bagging, the Boosting technique creates a variety of training data sets by random sampling with replacement from the primary training dataset. The sequence of models is trained in this method. Initially, the first model is created from the training dataset. Next, a new model is created in order to modify the previous model according to its performance. Lastly, all the weak learners are combined by weighted majority voting to build the final strong model (Schapire, 2003; Alfaro et al., 2013). The illustration of the Boosting algorithm process is shown in Fig. 4.

There are some most notable examples of this approach, including Stochastic Gradient Boosting (SGB), AdaBoost, and eXtreme Gradient

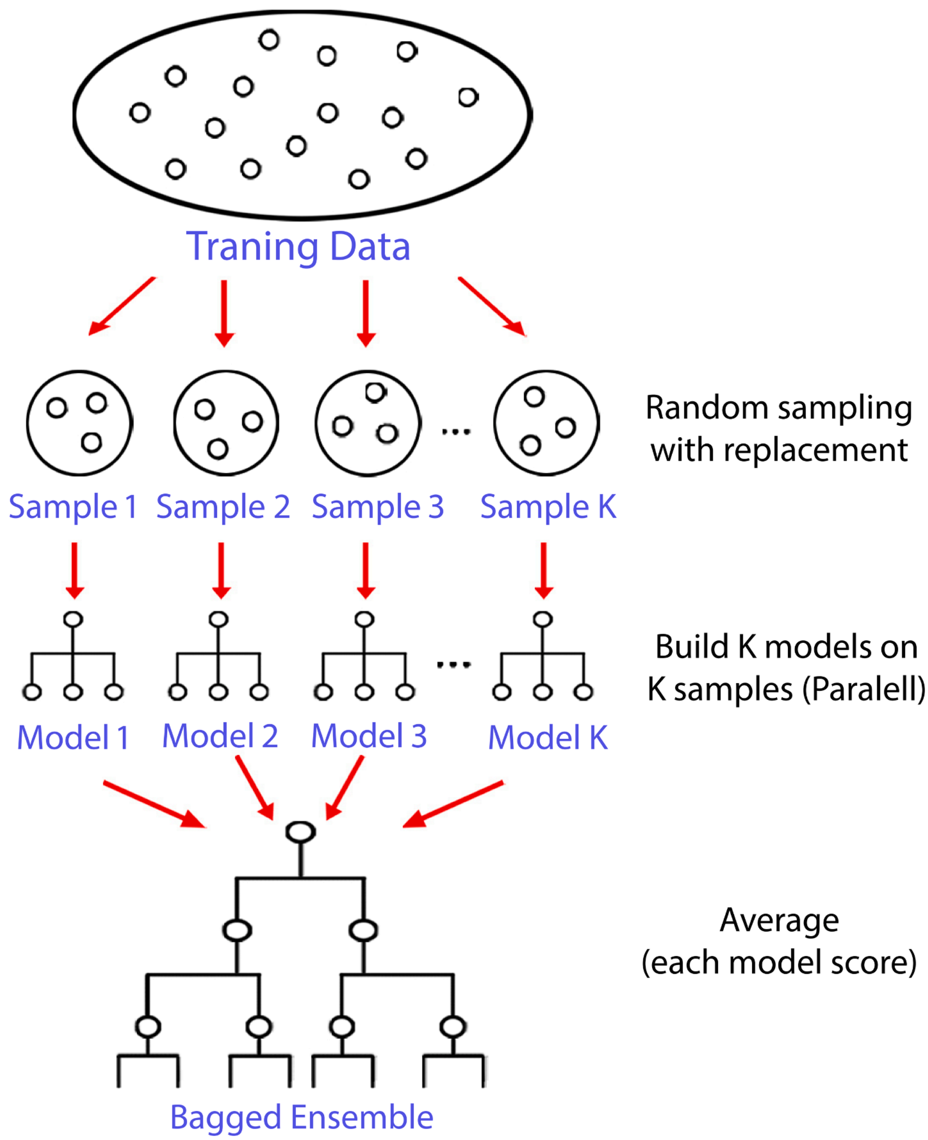

Fig. 3. Schematic illustration of the bagging ensemble learning for prediction problems.

Boosting (XGB), which are briefly described as follows.

3.4.1. AdaBoost

The AdaBoost (also called Adaptive Boosting) algorithm introduced by Freund and Schapire (1996) is one of the popular iterative boosting EML algorithms that applies effectively on a decision tree method. The AdaBoost technique converts weak learners into strong ones by overcoming their weaknesses and correcting the errors of weak learners by applying an iterative procedure. Weak learners are added in sequence and trained by the weighted training dataset. Then multiple weak learners are combined to create a strong one. The final strong output model is made by the weighted voting of weak learners (Freund and Schapire, 1997).

3.4.2. Stochastic Gradient Boosting

Stochastic Gradient Boosting (SGB) as introduced by Friedman (2002) is a machine learning method that is used for both regression and classification issues. Gradient Boosting creates a prediction model utilizing an ensemble of weak classifiers (e.g., decision trees). The SGB algorithm is a randomizing form of the standard Gradient Boosting algorithm.

In SBG, the sub-samples of the original training dataset are selected randomly without replacement, then these random sub-samples (rather than the whole dataset) are applied to fit the weak learners and estimate the new model for the current iteration. It eventually constructs the final prediction model by adding up the predictions (of all trees). Similar to the bootstrap sampling technique (in the Bagging method), a stochastic sub-sampling is able to make a reduction in the variance of gradient boosting approximations (Friedman, 2002).

3.4.3. Extreme Gradient Boosting

Extreme Gradient Boosting (XGB, also known as XGBoost) is another boosting method based on gradient boosting that uses more precise approximations to create the best prediction model (among gradient boosting models) (Friedman, 2001). The most obvious advantage of XGBoost, among other gradient boosting methods is its speed. In this technique, the decision trees classifier is commonly used as a weak model.

This method's basic function is to predict a new classification membership after each iteration and it is performed in an additive fashion. Thus, the predictions are created from weak learners that continuously develop over the mistakes of the former learners. Finally, similar to other EML methods, XGBoost creates the final prediction model based on a combination of all the previous individual models

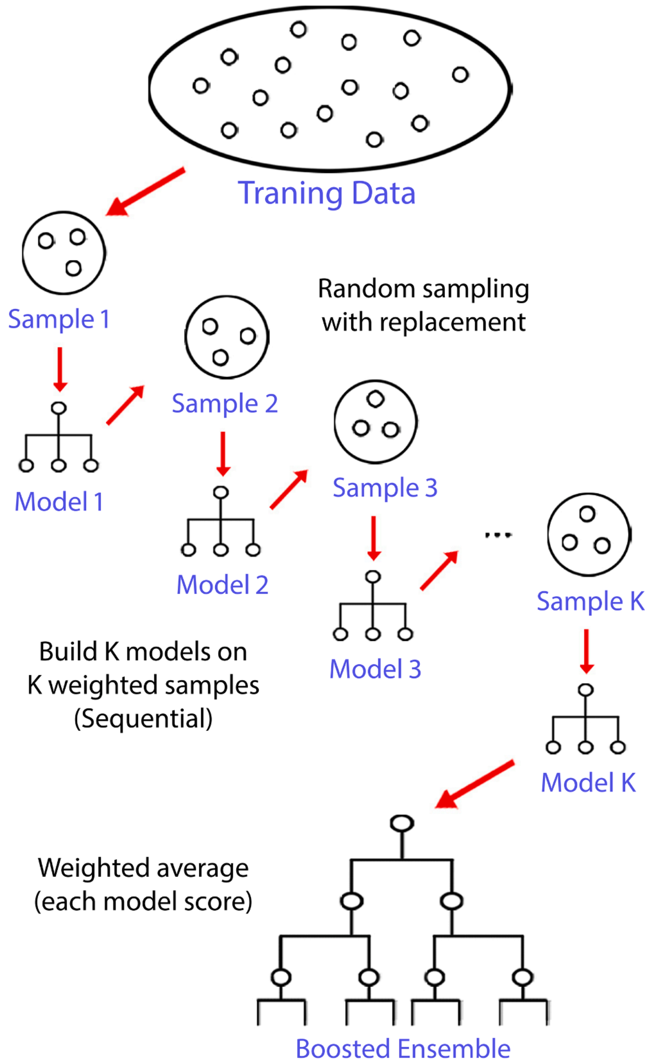

Fig. 4. Schematic illustration of the boosting ensemble learning for prediction problems.

(Friedman, 2001; Georganos et al., 2018).

3.5. Dagging

The dagging algorithm (also known as Disjoint Aggregating) is another ensemble machine learning method that is applied to build *meta*-learners (Ting and Witten, 1997). Dagging is very similar to bagging but has a different sampling method. In this method, rather than bootstrap sampling, the disjoint sampling method is used to obtain random training subsets from the original dataset without replacement (Yariyan et al., 2020). Finally, the majority vote technique is employed for combining the individual output models obtained from disjoint samples.

4. Ensemble learning in Hydrology, Review of the State-of-the-art

Hydrological systems involve complex processes, which makes it inadequate to employ linear (or even in some cases, simple nonlinear) statistical models to properly simulate or predict their properties. In the last two decades, several comparisons have been carried out between the simulative and predictive performance of classical methods, such as conceptual and stochastic methods (Sharma et al., 2019), and soft computing methods, such as neurocomputing methods and support vector machines (Lange and Sippel, 2020; Alizamir et al., 2020a, 2020b). In most studies, the better performance of soft computing methods over classical methods has been declared (Zhu et al., 2019). In short, nonlinear soft computing (e.g., machine-learning approaches) has shown to be successfully dealing with different aspects of hydrological systems (Kisi and Heddam, 2019; Zounemat-Kermani et al., 2020a). However, individual and single ML models might fail in providing promising results for the predictions or simulations of hydrological systems due to uncertainties in training parameters. In that sense, ensemble approaches were deemed as effective tools to mitigate the modeling errors (Zounemat-Kermani et al., 2020b) and overfitting problem (Berkhahn et al. 2019). Despite the existence of numerous studies on the use of soft computing techniques that aim to model hydrological processes, far less studies were developed to account for the implication of EML methods in hydrology. In the following subsections, detailed descriptions of past researches as well as the state-of-the-art of the recent developments and implications concerning the application of EML methods in different areas of hydrology, are presented.

4.1. Surface hydrology

In recent years, EML approaches have shown their potential in simulation/prediction of surface hydrological processes and could be considered as appropriate approaches for improving the performance of single ML models (Adnan et al., 2020a). In this section, the state-of-the-art and latest implications of ensemble approaches in three aspects of surface hydrological systems, including (i) river flow and streamflow modeling, (ii) surface hydrological processes, and (iii) rainfall-runoff simulation, are explored and investigated in details.

4.1.1. River flow and streamflow modeling

Streamflow and river flow processes are important key elements in water resources management, irrigation planning, water demand evaluation, hydroelectric generation, and flood warning systems. Although it has been stated that modeling river/stream systems - either for time series forecasting or quantitative predictions - is a demanding and difficult task to do, some researchers alleged that the employment of EML models can be considered as one of the most suitable approaches (Tyralis et al., 2020; Xu et al., 2020).

In the past two decades, several researchers have attempted to utilize the EML methods in modeling river- and streamflow in terms of flow prediction and forecasting. As one of the earliest studies, Cannon and Whitfield (2002) investigated the changes in streamflow due to the impacts of climate change using Ensemble Neural Network (ENN) and Multiple Linear Regression (MLR) methods. The outcomes of the models revealed the better modeling performance of the ENN model in comparison to the MLR ones (around 4% improvement in the coefficient of determination). Diks and Vrugt (2010) employed several model averaging EML methods for streamflow forecasting. The methods applied consist of BMA, EWA, BGA, AICA, BICA, MMA, and GRA. The findings of the study demonstrated the superiority of the GRA compared to the other applied averaging methods.

Li et al. (2010) used the Support Vector Machines (SVM) and MLR as weak learners for constructing an EML model based on the bagging methodology. The findings of the research denote that the ensemble methods, including the bagged-SVM and bagged-MLR outperformed the individual models. Tiwari and Chatterjee (2011) tried to forecast daily river discharge by the aid of bootstrap and wavelet Artificial Neural Networks (ANN). The results indicated that the hybrid wavelet bootstrapped ANN model provided the better predictions than individual ML models.

An appraisal of bagging Stochastic Gradient Boosting (SGB) ensemble methods in streamflow prediction was conducted by Erdal and Karakurt (2013). Tree-based base learners, namely the Classification And Regression Trees (CART), were employed for building the EML models such as Bagged Regression Trees (BRT) and Stochastic Gradient

Boosted Regression Trees (GBRT). It was found that the ensemble BRT and GBRT models were superior to individual methods of CART and Support Vector Regression (SVR) models (around 17% improvement in the predicted RMSE values). Kim et al. (2019) developed an ensemble learning regression ML method for estimating discharge using satellite altimetry data. By combining three traditional rating curves as base learners, the results of the model applied showed that the ensemble technique is an effective ML approach to predict the discharge in the central Congo River.

Galelli and Castelletti (2013) evaluate the accuracy and computational efficiency of ensemble Extremely Randomized Trees (ERT) in modeling streamflow. The applied ERT model surpassed the individual models such as CART, M5 model tree, ANN, and MLR (around 10% improvement in the modeled RMSE values). Tyralis et al. (2019) aimed to forecast short-term daily streamflow by the aid of the stacking strategy using ten different ML methods. The ensemble model acted better than the individual models. In another study, Li et al. (2020a) integrated several Data-Driven Models (DDMs) using a stacking ensemble approach for mid-term streamflow forecasting. The applied DDMs include SVR, Random Forest (RF), Elastic Net Regression (ENR), and eXtreme Gradient Boosting (XGB). It was found that the application of the stacking strategy improved the ability of individual models.

Zhou et al. (2018) compared the performance of three individual ANN models, such as Radial Basis Function Neural Network (RBFNN), the Elman network, and Extreme Learning Machine (ELM) versus three averaging ensemble techniques. It was found that the EML models attained better predictions in comparison to the individual ANNs. Table 1 represents a summary of the research papers using ensemble techniques in modeling river and streamflow. It can be seen from Table 1 that almost all of the earlier studies (until 2010) employed the bagging technique in their methodology. Afterward, researchers tended to use other EML methods such as boosting and stacking. Most recently, newly developed ensemble methods e.g., modified stacking, and XGB have been applied for streamflow forecasting (Yu et al., 2020; Li et al., 2020a, 2020b).

Table 1
Summary of the application of ensemble learning methods in river and streamflow modeling.

Researcher(s)	Motivation	Ensemble method(s) applied	EML approach/ Base learner(s)
Cannon and Whitfield (2002)	Streamflow forecasting	Bagging	ANN
Dakou et al. (2007)	River flow prediction	Bagging, Boosting	Tree-based
Boucher et al. (2010)	Streamflow forecasting	Bagging	ANN
Li et al. (2010)	Streamflow prediction	Bagging	SVM
Diks and Vrugt (2010)	Streamflow forecasting	Model averaging	EWA, BMA, BGA, AICA, BICA, MMA, GRA
Tiwari and Chatterjee (2011)	River discharge forecasting	Bagging	Wavelet-ANN
Galelli and Castelletti (2013)	Streamflow modelling	Bagging	ERT, Tree-based
Erdal and Karakurt (2013)	Monthly streamflow prediction	Bagging, oosting	SGB, Tree-based
Kim and Seo (2015)	Streamflow forecasting	Stacking	ANN
Arsenault et al. (2015)	Streamflow simulation	Model averaging	AICA, BGA, BMA, GRA, SAM
Najafi and Moradkhani (2016)	Seasonal streamflow forecasting	Bagging, Model averaging	MLR, BMA
Qu et al. (2017)	River flow forecasting	Model averaging	BMA
Zhou et al. (2018)	Monthly streamflow forecasting	Stacking	ANNs
Ghorbani et al. (2018)	River flow prediction	Model Averaging	ANN, SVM
Kim et al. (2019)	Estimating river discharge	Stacking	Statistical model
Tyralis et al. (2019)	Daily streamflow forecasting	Stacking	SVR, Tree-based Regressive
Yu et al. (2020)	Streamflow forecasting	Boosting	XGB, Tree-based
Li et al. (2020a)	Monthly streamflow forecasting	Boosting, Stacking, Bagging	XGB, Tree-based
Tyralis et al. (2020)	Daily streamflow forecasting	Boosting, Bagging	XGB, Tree-based
Xu et al. (2020)	Daily river discharge modeling	Boosting	Statistical model
Kim et al. (2020)	Daily streamflow forecasting	Model averaging	BMA, MARS, M5Tree, ELM
Ni et al. (2020)	Streamflow forecasting	Boosting	XGBoost
Darbandsari and Coulibaly (2020)	Daily streamflow forecasting	Model averaging	BMA
Alobaidi et al. (2021)	Seasonal low-flow modeling	Stacking	ANN
Sihag et al. (2021)	River basin discharge estimation	Bagging	RF

4.1.2. Surface Hydrological Processes

In general, surface hydrological processes involve a variety of quantitative and qualitative surface water systems that are significant components in water resources management and engineering. In addition to the descriptions given in the previous section regarding the streamflow/river flow, in the light of the present literature, ensemble approaches have been successfully used concerning other areas of surface hydrological processes. Instances include modeling of river water quality parameters, debris flow, river icing, and suspended sediment concentration. In this regard, some researchers attempted to use ensemble techniques in modeling water quality in surface water bodies such as river flow (Napiorkowski et al., 2014; Shin et al., 2017; Elkiran et al., 2019). Barzegar et al. (2018b) explored the potential of boosting ensemble technique in forecasting electrical conductivity (EC) as a water quality parameter, using ELM models as base learners. The results showed that boosting the ELM model outperformed the individual ML models. Lu and Ma (2020) applied two types of ensemble methods i.e., XGB and modified bagging (RF model), for predicting short-term water quality. Six water quality parameters (specific conductance, pH value, water temperature, dissolved oxygen, fluorescent dissolved organic matter, and turbidity) were considered for the modeling process. The obtained results demonstrated that the applied RF model acted best in the prediction of dissolved oxygen, temperature, and specific conductance.

Other scholars applied EML for modeling surface water hydrological processes, such as debris flow (Mei et al., 2018), river icing (Barzegar et al., 2019), and suspended sediment concentration (Alizadeh et al., 2017). Pai et al. (2014) applied the AdaBoost ensemble method for analyzing debris resulting from natural disasters such as typhoons and heavy rainfalls. It was claimed that better prediction results were attained by the usage of the ensemble technique in comparison to the previous investigations (more than 7% improvement in the accuracy). Shamaei and Kaedi (2016) employed the stacking method for improving the suspended sediment prediction results of individual ML models including Linear Genetic Programming (LGP) and Neuro-Fuzzy (NF). The results revealed that the applied ensemble method improved the performance of the individual models (around 40% improvement in the RMSE values). Sun and Trevor (2018) applied the stacking ensemble learning to forecast breakup dates of river ice. They used Bayesian regularization ANN and ANFIS as individual and combined ensemble model based on the simple average method. It was concluded that combining models outperformed the individual models in terms of average squared errors by around 16%.

Table 2 summarizes the studies corresponding to employing EML models in surface hydrology processes. According to the researches applied in Table 2, it can be observed that different ensemble models (bagging, boosting, and stacking) have been employed for various surface hydrology purposes.

4.1.3. Rainfall-runoff Process

In hydrology, accurate rainfall-runoff modeling is an important research topic with impact on subsequent water resources planning, including dam engineering, water distribution plans, watershed management, and flood control (Adnan et al., 2020b). Although there are a lot of publications on the application of ML models and rainfall-runoff process, there have been only a few published studies focused on the usage of EML in modeling rainfall-runoff. Instances include the early works of Abrahart (2003) and Jeong and Kim (2005) who successfully applied bootstrap and bagging techniques on ANN base-learners. In Jeong and Kim's (2005) study, two types of individual and bagging ANNs have been used to simulate the rainfall-runoff process. The findings of the study revealed the superiority of the ensemble model in comparison to the other applied ones.

Sharma and Tiwari (2009) undertook a study to predict monthly rainfall-runoff using ensemble bootstrap ANN. The results indicated that considering and combining soil, geomorphology, topography, and vegetation inputs in the model lead to a better prediction for the applied ensemble model (around 28% improvement in the RMSE values). Liu et al. (2014) applied the AdaBoost (Adaptive Boosting) ensemble technique for the first time for efficiency improvement of rainfall–runoff models. It was reported that the improved AdaBoost model provided more acceptable results (4% improvement in coefficient of determination). Recently, Bennett (2019) used the XGB technique for the uncertainty analysis and parameter identification of a rainfall-runoff model. They claimed that the use of the XGB in training surrogate models offered an efficient outline for inferring the model parameter and uncertainty analysis.

Table 3 presents the summary of the application of EML for modeling the rainfall -runoff processes. It is obvious that the majority of the applied models got the advantage of the ANNs as for the base learners ML models.

4.2. Hydrogeology, Groundwater Modeling

Unlike surface hydrological processes, the main challenge in hydrogeology and flow in porous media is to deal with a shortage of data on soil and aquifer properties. Thus, the application of capable nonlinear models that can capture the complex behaviors of hydrogeological problems, such as predicting the position of the groundwater table or the fate of a contaminant in the subsurface, is of great importance in this field. In this section, it is intended to address the past and recent developments of ensemble methods in quantitative and qualitative groundwater modeling.

Laucelli et al. (2007) applied a simple averaging method based on symbolic regression for obtaining more reliable approximations in

Table 2
Summary of the application of ensemble learning methods in dealing with surface hydrological processes.

General Topic	Research	Motivation	Ensemble methods applied	EML approach/Base learner(s)
River water quality	Napiorkowski et al. (2014)	Stream temperature forecasting	Bagging	ANN
	Shin et al. (2017)	Predicting cyanobacterial blooms in river	Bagging, Boosting	AdaBoost, Tree-based
	Barzegar et al. (2018b)	Electrical conductivity (EC) forecasting	Boosting	WA-ELM
	Elkiran et al. (2019)	Predicting River water quality	Model averaging	ANN, ANFIS, SVM
	Kisi et al. (2020)	Prediction of dissolved oxygen	Model averaging	BMA, ANN, ANFIS, CART, MLR ELM
	Lu and Ma (2020)	Predicting six water quality indicators	Boosting,Bagging	XGB, Tree-based
	Abba et al. (2020b)	prediction of dissolved oxygen concentration	Model averaging	SAM, LSTM, ELM, GRNN
	Lu and Ma (2020)	Short-term water quality prediction	Boosting	XGB, Tree-based
	Sharafati et al. (2020a)	Effluent quality parameters	Boosting, Bagging	AdaBoost, GBRT, RF, Regression-based
	Melesse et al. (2020)	River water salinity modeling	Bagging	RF, RS, M5model
	Abba et al., 2020a	Predicting water quality index	Boosting	XGB
	Abba et al. (2020c)	Prediction of water quality index	Stacking	ANN, ANFIS, SVR, MLR
	Alizamir et al. (2021)	Estimating daily chlorophyll-a concentration	Bagging	RF
	Heddam (2021a), Heddam (2021b)	Predicting DO Concentration in River	Bagging	ERT, RF
	Pai et al. (2014)	Debris flow	Boosting	AdaBoost, ANN
Debris flow	Mei et al. (2018)	Debris flow forecasting	Bagging	Tree-based
	Sun and Trevor (2018)	River ice breakup dates modeling	Stacking	ANN, ANFIS
River icing	Barzegar et al. (2019)	Estimate river ice thickness	Bagging	ELM, SSVM
	Sun et al. (2020)	River ice breakup dates modeling	Stacking	Statistical models
	Francke et al. (2008)	Estimation of suspended sediment concentration and	Bagging	Tree-based
Suspended sediment transportation	Oehler et al. (2012)	Suspended sediment reference concentration	Boosting	Tree-based
	Singh and Panda (2015)	Daily sediment estimation from a small agricultural watershed	Bootstrap	ANNs
	Wang et al. (2015)	Sediment yield modeling	Model averaging	BMA
	Shamaei and Kaedi (2016)	Suspended sediment concentration modeling	Stacking	GP & ANFIS
	Alizadeh et al. (2017)	Forecasting suspended sediment concentration	Stacking	WA-ANNs
	Nhu et al. (2020a), Nhu et al. (2020b)	Monthly suspended sediment load prediction	Bagging	RF, RSM, Tree-based
	Sharafati et al. (2020c)	Daily suspended sediment load prediction	Bagging, Boosting	RF, AdaBoost, Tree-based
	Shadkani et al. (2020)	Predicting daily suspended sediment load	Boosting	GBRT
	Moeini et al. (2021)	Estimation of Total Suspended Solids	Boosting, Bagging	RF, AdaBoost, Tree-based

Table 3
Summary of the application of ensemble learning methods in modeling rainfall-runoff processes.

Research	Motivation	Ensemble methods applied	EML approach/ Base learner(s)
Abrahart (2003)	Rainfall-runoff forecasting	Bagging	ANN
Jeong and Kim (2005)	Rainfall-runoff modelling	Bagging	ANN
Sharma and Tiwari (2009)	Monthly runoff prediction	Bagging	ANN
Dong et al. (2013)	Rainfall-runoff simulation	Model averaging	BMA
Phukoetphim and Shamseldin (2013)	Rainfall-runoff simulation	Boosting	Stochastic GB
Liu et al. (2014)	rainfall–runoff forecast	Boosting	AdaBoost, Regression-based
Kan et al. (2015)	Event-based rainfall-runoff simulation	Bagging	ANN
Bennett (2019)	Uncertainty analysis of rainfall-runoff	Boosting	Regression-based
Huang et al. (2019)	Monthly runoff prediction	Model averaging	BMA
He et al. (2020)	Monthly runoff forecasting	Boosting	Tree-based

modeling the groundwater balance. It was concluded that applying the ensemble method was a reasonable strategy to reduce simulation errors due to variance. Singh et al. (2014) employed two types of ensemble tree-based methods, such as the decision tree forest and decision tree-boost for modeling groundwater hydrochemistry in northern regions of India. The findings of the study revealed that the applied ensemble models acted better than the single decision tree model, and were successful in delineating the influences of anthropogenic activities and seasonal variations on groundwater hydrochemistry.

Barzegar et al., (2017) assessed the efficiency of different wavelet-based ML models, including wavelet Extreme Learning Machine (ELM), wavelet Group Method of Data Handling (GMDH), and wavelet ANN in forecasting the groundwater level. In comparison to the advanced applied ML models, it was claimed that the boosting ensemble models surpassed the individual ML models (more than 50% improvement in the RMSE values). In another study, Barzegar et al. (2018a), Barzegar et al. (2018b) used ensemble modeling for mapping groundwater contamination risks in multiple aquifers. The outcomes of the study revealed that combining the results of different individual ANN models yielded higher accuracy than the individual ML models such as SVR and MARS.

Naghibi et al. (2019) applied several EML methods, including rotation forest, boosted regression tree, and random forest techniques for modeling groundwater potential maps. They concluded that an improved EML based on the evidential belief function gave a better performance in comparison to the other applied ensemble models and individual CART model. Avand et al. (2020) compared the performance of several EML models i.e., MultiBoosting, Best-First tree, Bagging, and AdaBoost also for groundwater potential mapping and influence of the effective parameters. It was shown that among the applied ensemble techniques, the Best-First tree algorithm had the highest simulation accuracy, with an average of 4% improvement based on the classification accuracy. In a comprehensive study, Chen et al. (2020a), Chen et al. (2020b) applied different types of tree-based ensemble learning methods, such as AdaBoost, bagging, dagging, RSM, and RtF in again groundwater potential mapping. The performance of the applied techniques was assessed, and the outcomes showed that all the EML techniques illustrated a good predictive performance; nevertheless, the rotation forest gave the best modeling results.

In Table 4, a concise report of the applied EML techniques in groundwater modeling is presented. Likewise, the popular implementation of ensemble learning in surface hydrology, the use of EML methods in hydrogeology, especially in groundwater potential mapping, has attracted the scholars' attention, specifically in the last couple of years (see Table 4). It can be observed that there are several pertinent studies to the year 2020. Needless to say, in addition to the conventional bagging and boosting methods, the AdaBoost EML technique has also been mostly considered in the recent works.

Table 4
Summary of the application of ensemble learning methods in modeling groundwater processes and potential mapping.

Research	Motivation	Ensemble methods applied	EML approach/Base learner(s)
Laucelli et al. (2007)	Groundwater balance	Model averaging	Regression-based
Rojas et al. (2008)	Groundwater modeling	Model averaging	BMA
Li and Tsai (2009)	Groundwater head prediction	Model averaging	BMA
Rodriguez-Galiano et al. (2014)	Groundwater nitrate pollution	Bagging	RF
Singh et al. (2014)	Groundwater hydrochemistry modeling	Bagging, Boosting	Tree-based
Chitsazan and Tsai (2015)	Groundwater prediction	Model averaging	BMA
Nolan et al. (2015)	Simulating groundwater nitrate models	Boosting	BRT
Barzegar et al. (2017)	Groundwater level fluctuations	Boosting	GMDH, ELM
Liu and Merwade (2018)	Flood inundating modeling	Model averaging	BMA
Barzegar et al. (2018a)	Mapping groundwater contamination	Model averaging	ANN
Naghibi et al. (2019)	Spatial groundwater potential mapping	Boosting	Regression tree
Rizeei et al. (2019)	Groundwater aquifer potential modeling	Boosting	ANN, SVM
Avand et al. (2020)	Groundwater potential mapping	Boosting Bagging AdaBoost Best-First tree	Tree-based models
Nhu et al. (2020a), Nhu et al. (2020b)	Groundwater spring potential mapping	Stacking	Logistic regression
Chen et al. (2020b)	Groundwater spring potential mapping	AdaBoost, Bagging, RSM, Dagging, RtF	Tree-based
Nguyen et al. (2020)	Groundwater potential modeling	Bagging, Dagging, Decorate, Multiboost, RSM	Tree-based
Sharafati et al. (2020b)	Predicting groundwater level	Boosting	GBR
Sahour et al. (2020)	mapping the spatial distribution of groundwater salinity	Boosting	XGB
Friedel et al. (2020)	Predicting groundwater redox status	Bagging	RF, BRT
Naghibi et al. (2020)	Assessing groundwater spring potential	Bagging, Boosting	RF, XGB
Yin et al. (2021)	Groundwater storage change prediction	Model averaging	BMA

4.3. Extreme Hydrological Events

Appropriate modeling in terms of estimation of the occurrence frequency, prediction, and forecasting of hydrological extreme events, such as floods and droughts, is of high importance for the design, public safety, operation, and maintenance of hydraulic and water resources systems. Hence, this section tries to explore the utilization of ensemble modeling in extreme hydrological events, including flood susceptibility mapping, flood modeling, and drought forecasting.

4.3.1. Flood susceptibility mapping and flood modeling

Floods are categorized as one of the most hazardous environmental events and natural disasters, causing immense economic, public, social, and environmental damages worldwide (Wang et al., 2019). Having promising hydrological prediction and forecasting models would lead to establishing adequate flood warning systems that are applicable tools for assessing flood risk and reducing consequent damages by forecasting flood events. Moreover, flood susceptibility mapping can lead to the reduction of flood damages costs in terms of predicting future flood occurrences. Some researchers have already tried to employ ensemble learning methods in flood modeling, such as flood prediction (Li et al., 2013; Venkatesan and Mahindrakar, 2019) and geospatial mapping of flood susceptibility (Arabameri et al., 2020; Pham et al., 2021). Shu and Burn (2004) estimated the index flood using an incorporated boosting and bagging ensemble techniques by the aid of ANNs as the weak learners. It was shown that the ensemble ANNs were considerably more accurate than the individual ANN models (around 10% improvement in relative squared error). Tiwari and Chatterjee (2010) successfully created a bootstrap model based on ANNs and wavelet decomposition technique for hourly forecasting of flood.

In another study, Araghinejad et al. (2011) aimed to build an ensemble ANN system for flood forecasting. Having said that, the ensemble ANN enhance the accuracy of forecasted values up to 50% based on the calculated RMSE values. They also claimed that using an ensemble modeling system would improve the forecasting results. Lee et al. (2017) applied the boosted-tree models and the random-forest to map the flooded area in Seoul City, Korea. Both EML models acted appropriately in mapping the flood; however, the random forest ensemble model provided slightly better outcomes than the boosted-tree models.

Shafizadeh-Moghadam et al., (2018) implemented for flood susceptibility assessment several statistical and machine learning models, as well as seven ensemble approaches. Their findings revealed that the boosted regression trees model was superior to the individual machine learning models. That said, the boosted regression tree model surpassed the original ANN with 5% improvement in general performance. In a similar study, Arabameri et al. (2020) claimed that using an ensemble learning approach brought excellent accuracy in modeling flash flood susceptibility mapping. Table 5 summarizes the relevant applications of different EML methods in modeling flood and flood susceptibility mapping. It is obvious from Table 5 that the majority of the ensemble applications have focused on the usage of bagging techniques followed by the boosting methods.

Table 5
Summary of the application of ensemble learning methods in flood mapping and modeling.

Research	Motivation	Ensemble methods applied	EML approach/ Base learner(s)
Shu and Burn (2004)	Estimating the index flood	Bagging Boosting	ANN
Tiwari and Chatterjee (2010)	Flood forecasting	Bagging	Wavelet-ANN
Araghinejad et al. (2011)	Flood forecasting	Stacking	ANN
Liang et al. (2013)	Flood forecasting	Model averaging	BMA
Li et al. (2016)	Flood forecasting	Boosting	SVM
Chapi et al. (2017)	Flood susceptibility assessment	Bagging	Tree-based
Lee et al. (2017)	Prediction of flood susceptibility	Bagging Boosting	RF, CART
Al-Abadi (2018)	Mapping flood susceptibility	Bagging, Boosting	RF, RtF, AdaBoost
Shafizadeh-Moghadam et al. (2018)	Flood susceptibility mapping	Boosting	Regression tree
Venkatesan and Mahindrakar (2019)	Flood forecasting	Boosting	XGB, Tree-based
Chen et al. (2019)	Flood susceptibility modelling	Bagging	RSM, Tree-based
Bui et al., 2019	Flash flood susceptibility modeling	Bagging, Boosting	AdaBoost, Fuzzy based algorithm
Mosavi et al. (2020)	Flood and erosion susceptibility mapping	Model averaging	Linear model, MARS, RF
Chen et al. (2020a)	Flood risk assessment	Bagging	Tree-based
Hosseini et al. (2020)	Flash-flood hazard assessment	Boosting, Bagging (RF), BMA	Tree-based
Yariyan et al. (2020)	Flood probability mapping	Bagging Dagging	Tree-based
Islam et al. (2020)	Flood susceptibility modelling	Dagging, Bagging	RSM, ANN, RF, SVM
Arabameri et al. (2020)	Flash flood susceptibility modeling	Bagging	Tree-based
Pham et al. (2021)	Flood susceptibility mapping	Bagging	RSM, Tree-based
Mirzaei et al. (2021)	Flood susceptibility assessment	Boosting	XGB

4.3.2. Drought Prediction and Forecasting

Drought is an important intense hydrological event; its prediction and forecasting are of critical importance in water resources planning and management. Frequent studies have already applied statistical (Schick et al., 2018) and dynamical (Turco et al., 2017) methods in predicting drought. However, being a highly unpredictable phenomenon, several researchers have made efforts to use conventional and recent ML models for the sake of drought modeling (Kisi et al., 2019). However, to the best of the authors' knowledge, ensemble learning methods have not been employed extensively in modeling hydrological drought. In the following, a few available applications of EML in drought modeling are mentioned. Belayneh et al. (2016) investigated the potential of coupled ensemble learning approaches in predicting drought in Ethiopia. The bootstrap and boosting ensemble techniques were established based on ANN and SVR weak learners. It was reported that the boosting EML technique improved the prediction results in comparison to the individual models.

Xu et al. (2018) developed several statistical and machine learning models for the prediction of drought in China. They used the ensemble model averaging for improving the results of their models. It was concluded that multi-model ensembles had considerable advantages over single dynamic models. Li et al. (2020b) applied ensemble learning techniques for predicting drought conditions. They used regression and decision tree models for the weak learners to build the Adaboost and RF ensemble models. It was reported that the ensemble methods were superior to their individual counterpart tree-based models. Specifically, the RF model surpassed the DT model around 45% improvement according to the RMSE values. In Table 6, the report of the above-mentioned applications is presented.

5. Discussion

From the onset of the emergence and application of ML and SC techniques in hydrology, numerous studies have been conducted and reported in the literature (refer to Tables 1 to 6). Nonetheless, the majority of these studies are structured based on the regular learning of the machine learning models (Zounemat-Kermani et al., 2019). Ensemble methods try to seek several potential solutions before representing the ultimate prediction/simulation of a dynamic system. In short, such

Table 6
Summary of the application of ensemble learning methods in modeling drought.

Research	Motivation	Ensemble methods applied	EML approach/ Base learner(s)
Belayneh et al. (2016)	Drought prediction	Bootstrap, Boosting	ANN, SVR
Xu et al. (2018)	One-six month lead drought prediction	Model averaging	SVRs
Zhai et al. (2020)	Drought characteristics prediction	Model averaging	SAM
Li et al. (2020b)	Drought forecasting	Boosting, Bagging	AdaBoost, Tree-based
Buthelezi et al. (2020)	Monitoring drought damage	Bagging, Boosting	XGB, RF

learning approaches are to weigh several data-driven individual models (e.g., machine learning, statistical, & stochastic), and combine the results of them in order to advance the predictive/simulative performance of the individual (base learners) models.

In this review, we explored and presented studies that aim at modeling different areas of surface hydrology, river water quality, groundwater modeling, and hydrological events using ensemble paradigm with distinguished strategies (e.g., bagging, boosting, & stacking). Despite the fact that ensemble methods have not been applied widely in hydrological sciences, their applications have increased in recent years. Fig. 5 clearly illustrates the rapidly increasing trend of using EML methods in the last seven years (from 2014 to 2020) and also the great attention they have attained in the year 2020 alone. This fact becomes more interesting because one cannot expect the same ascending pace for the general use of neurocomputing models in hydrological sciences. A comprehensive review (from 2000 to 2019) carried out by Zounemat-Kermani et al. (2020b) shows that the application of neurocomputing models in hydrology follows a constant ascending trend from 2000, whereas EML approaches have gained attention from 2014.

Analyzing the ensemble learning researches in the areas related to hydrology shows that the most implemented base learners (42% of all the applied models) are based on the tree model structures (Fig. 6). The tree-based models, especially the regular classification and regression trees, are popular tools compared to other tree-based models such as the M5 model tree and CHAID techniques in hydrological modeling. Other EML tree-based models, including extra-trees are valid alternatives to traditional individual machine learning (e.g., ANN) and non-ensemble data-driven models (e.g., regression trees) approaches due to the fact that they can deliver efficiency equivalent (or higher) to those achievable with other common hydrological methods (Galelli and Castelletti, 2013). Among all the network-based machine learning models, such as ANNs, neuro-fuzzy models, gene expression programming, the ANNs are still widely applied as base learners for ensemble learning. As can be seen in Fig. 6, artificial neural networks hold the second place as popular base learners in constructing ensemble models (about 30%).

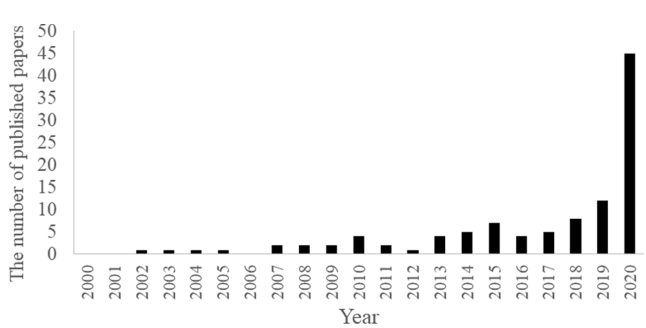

Fig. 5. Historical illustration of the application of ensemble methods in hydrology from 2000 to 2020 based on the Scopus database (the number of cases is in accordance with the mentioned references in Tables 1 to 6).

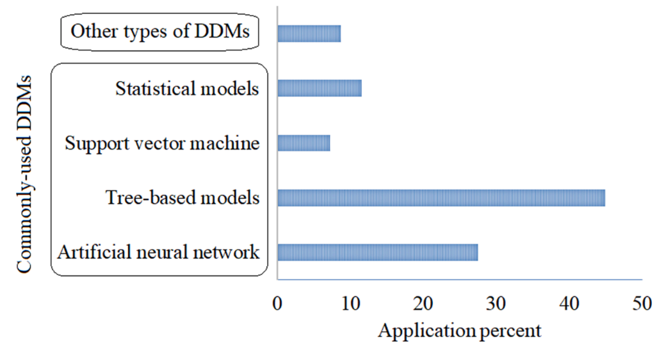

Fig. 6. The percentage application of different categories of base-learners in forming EML based on the Scopus database (from 2000 to 2020; the analysis is in accordance with the mentioned references in Tables 1 to 6).

Investigating the published researches implies that the majority of EML studies have mostly focused on the different versions of the bagging EML methods (mostly the regular bagging and random forest methods), followed by the boosting techniques (e.g., regular boosting, AdaBoost, and XGB). It is worth noting that most of the recent studies have a tendency to utilize more advanced ensemble methods such as Adaboost, XGB, and Dagging. In Fig. 7, the contribution of different ensemble categories in terms of percent application is illustrated.

It should be noted that in rare cases, the effect of ensemble techniques is reported to be inefficient. Dakou et al. 2007 claimed that the predictive performance of bagging and boosting (for predicting macroinvertebrate taxa in a river located in northern Greece) in some models was stable or even worsened. Overall, it is concluded that different approaches of ensemble learning paradigms can remarkably enhance the prediction/simulation accuracy of base learners and individual models in various fields of hydrology. For instance, Erdal and Karakurt (2013) reported that the CARTs gave promising results on monthly streamflow forecasting and yielded better performance than the support vector regression model.

6. Conclusions and Future Direction

Along with different machine learning approaches, EML models can successfully be employed for any hydrological problem and should be considered as the main candidate for complicated hydrological problems. There have been several available published cases for ensemble modeling in different areas of hydrology in recent years; nevertheless, there is limited literature on the application of well-known and novel machine learning models, such as ANFIS, GMDH, GEP, deep echo state, and ELM, as base learners in ensemble modeling. These individual models have proved their efficiency in dealing with complex problems in hydrology.

A general finding of this review implies that in the last three years (2018 to 2020), most researchers have aimed to apply novel EML methods including the AdaBoost, XGB, rotation forest, random

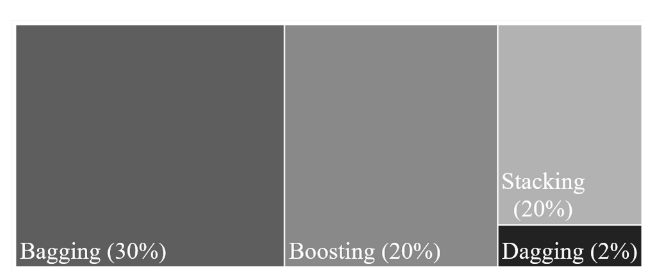

Fig. 7. The percentage applications of distinguished ensemble methods in hydrology based on the Scopus database (from 2000 to 2020; the analysis is in accordance with the mentioned references in Tables 1 to 6).

subspace, and dagging. Although the AdaBoost and XGB, as boosting methods, have been employed in several studies, further work is required to evaluate the ability of rotation and random subspace, as bagging methods, and dagging in different hydrological modeling. Another promising research direction is the integration of different ensemble strategies to improve a learning algorithm's performance. For instance, the stacking method can be combined with the bootstrap replicates to achieve further improvement in bagging/boosting methods.

In this review, several topics of hydrological sciences, including surface hydrology, hydrogeology, and hydrological events, were considered in evaluating the orientation of EML in hydrology. A broad evaluation of subjects related to hydraulics, hydrometeorology, and environmental factors that influence the hydrological cycle is beyond the scope of this paper. However, these topics could be the subject of future studies, and it is recommended that the importance of ensemble learning in other aspects of hydrological sciences should be appraised.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

As the corresponding author, I would like to express my gratitude to the Alexander von Humboldt Foundation for providing financial support for this research project within the framework of the Return Fellowship program.

References

Abba, S.I., Hadi, S.J., Sammen, S.S., Salih, S.Q., Abdulkadir, R.A., Pham, Q.B., Yaseen, Z. M., 2020a. Evolutionary computational intelligence algorithm coupled with self-tuning predictive model for water quality index determination. J. Hydrol. 587, 124974.

Abba, S.I., Linh, N.T.T., Abdullahi, J., Ali, S.I.A., Pham, Q.B., Abdulkadir, R.A., Anh, D.T., 2020b. Hybrid machine learning ensemble techniques for modeling dissolved oxygen concentration. IEEE Access 8, 157218–157237.

Abba, S.I., Pham, Q.B., Saini, G., Linh, N.T.T., Ahmed, A.N., Mohajane, M., Bach, Q.V., 2020c. Implementation of data intelligence models coupled with ensemble machine learning for prediction of water quality index. Environ. Sci. Pollut. Res. 27 (33), 41524–41539.

Abba, S.I., Nourani, V., Elkiran, G., 2019. Multi-parametric modeling of water treatment plant using AI-based non-linear ensemble. Journal of Water Supply: Research and Technology-Aqua 68 (7), 547–561.

Abrahart, R.J., 2003. Neural network rainfall-runoff forecasting based on continuous resampling. J. Hydroinf. 5 (1), 51–61.

Adnan, R.M., Liang, Z., Heddam, S., Zounemat-Kermani, M., Kisi, O., Li, B., 2020a. Least square support vector machine and multivariate adaptive regression splines for streamflow prediction in mountainous basin using hydro-meteorological data as inputs. J. Hydrol. 586, 124371.

Adnan, R.M., Petroselli, A., Heddam, S., Santos, C.A.G., Kisi, O., 2020b. Short term rainfall-runoff modelling using several machine learning methods and a conceptual event-based model. Stoch. Env. Res. Risk Assess. 1–20.

Al-Abadi, A.M., 2018. Mapping flood susceptibility in an arid region of southern Iraq using ensemble machine learning classifiers: a comparative study. Arabian J. Geosci. 11 (9), 218.

Alfaro, E., Gámez, M., Garcia, N., 2013. adabag: An R package for classification with boosting and bagging. J. Stat. Softw. 54 (2), 1–35.

Alizadeh, M.J., Nodoushan, E.J., Kalarestaghi, N., Chau, K.W., 2017. Toward multi-day-ahead forecasting of suspended sediment concentration using ensemble models. Environ. Sci. Pollut. Res. 24 (36), 28017–28025.

Alizamir, M., Heddam, S., Kim, S., Mehr, A.D., 2021. On the implementation of a novel data-intelligence model based on extreme learning machine optimized by bat algorithm for estimating daily chlorophyll-a concentration: case studies of river and lake in USA. J. Cleaner Prod. 285, 124868.

Alizamir, M., Kim, S., Kisi, O., Zounemat-Kermani, M., 2020a. A comparative study of several machine learning based non-linear regression methods in estimating solar radiation: Case studies of the USA and Turkey regions. Energy 117239.

Alizamir, M., Kim, S., Zounemat-Kermani, M., Heddam, S., Kim, N.W., Singh, V.P., 2020b. Kernel Extreme Learning Machine: An Efficient Model for Estimating Daily Dew Point Temperature Using Weather Data. Water 12 (9), 2600.

Alobaidi, M.H., Ouarda, T.B., Marpu, P.R., Chebana, F., 2021. Diversity-driven ANN-based ensemble framework for seasonal low-flow analysis at ungauged sites. Adv. Water Resour. 147, 103814.

Arabameri, A., Saha, S., Chen, W., Roy, J., Pradhan, B., Bui, D.T., 2020. Flash flood susceptibility modelling using functional tree and hybrid ensemble techniques. J. Hydrol. 587, 125007.

Araghinejad, S., Azmi, M., Kholghi, M., 2011. Application of artificial neural network ensembles in probabilistic hydrological forecasting. J. Hydrol. 407 (1–4), 94–104.

Arsenault, R., Gatien, P., Renaud, B., Brissette, F., Martel, J.L., 2015. A comparative analysis of 9 multi-model averaging approaches in hydrological continuous streamflow simulation. J. Hydrol. 529, 754–767.

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. Int. J. Digital Earth 1–22.

Barzegar, R., Fijani, E., Moghaddam, A.A., Tziritis, E., 2017. Forecasting of groundwater level fluctuations using ensemble hybrid multi-wavelet neural network-based models. Sci. Total Environ. 599, 20–31.

Barzegar, R., Ghasri, M., Qi, Z., Quilty, J., Adamowski, J., 2019. Using bootstrap ELM and LSSVM models to estimate river ice thickness in the Mackenzie River Basin in the Northwest Territories. Canada. *Journal of Hydrology* 577, 123903.

Barzegar, R., Moghaddam, A.A., Adamowski, J., Ozga-Zielinski, B., 2018a. Multi-step water quality forecasting using a boosting ensemble multi-wavelet extreme learning machine model. Stoch. Env. Res. Risk Assess. 32 (3), 799–813.

Barzegar, R., Moghaddam, A.A., Deo, R., Fijani, E., Tziritis, E., 2018b. Mapping groundwater contamination risk of multiple aquifers using multi-model ensemble of machine learning algorithms. Sci. Total Environ. 621, 697–712.

Belayneh, A., Adamowski, J., Khalil, B., Quilty, J., 2016. Coupling machine learning methods with wavelet transforms and the bootstrap and boosting ensemble approaches for drought prediction. Atmos. Res. 172, 37–47.

Bennett, F.R., 2019. Gradient boosting machine assisted approximate Bayesian computation for uncertainty analysis of rainfall-runoff model parameters. 23rd International Congress on Modelling and Simulation.

Berkhahn, S., Fuchs, L., Neuweiler, I., 2019. An ensemble neural network model for real-time prediction of urban floods. J. Hydrol. 575, 743–754.

Boucher, M.A., Laliberté, J.P., Anctil, F., 2010. An experiment on the evolution of an ensemble of neural networks for streamflow forecasting. Hydrol. Earth Syst. Sci. 14 (3), 603–612.

Breiman, L., 1994. Bagging predictors (Technical Report 421). University of California, Berkeley.

Breiman, L., 2001. Random forests. Machine learning 45 (1), 5–32.

Bui, D.T., Tsangaratos, P., Ngo, P.T.T., Pham, T.D., Pham, B.T., 2019. Flash flood susceptibility modeling using an optimized fuzzy rule based feature selection technique and tree based ensemble methods. Sci. Total Environ. 668, 1038–1054.

Buthelezi, M.N.M., Lottering, R.T., Hlatshwayo, S.T., Peerbhay, K., 2020. Comparing rotation forests and extreme gradient boosting for monitoring drought damage on KwaZulu-Natal commercial forests. Geocarto International 1–24.

Cannon, A.J., Whitfield, P.H., 2002. Downscaling recent streamflow conditions in British Columbia, Canada using ensemble neural network models. J. Hydrol. 259 (1–4), 136–151.

Chapi, K., Singh, V.P., Shirzadi, A., Shahabi, H., Bui, D.T., Pham, B.T., Khosravi, K., 2017. A novel hybrid artificial intelligence approach for flood susceptibility assessment. Environ. Modell. Software 95, 229–245.

Chen, J., Li, Q., Wang, H., Deng, M., 2020a. A machine learning ensemble approach based on random forest and radial basis function neural network for risk evaluation of regional flood disaster: a case study of the Yangtze River Delta, China. Int. J. Environ. Res. Public Health 17 (1), 49.

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. J. Hydrol. 575, 864–873.

Chen, W., Zhao, X., Tsangaratos, P., Shahabi, H., Ilia, I., Xue, W., Ahmad, B.B., 2020b. Evaluating the usage of tree-based ensemble methods in groundwater spring potential mapping. J. Hydrol. 583, 124602.

Chitsazan, N., Tsai, F.T.C., 2015. A hierarchical Bayesian model averaging framework for groundwater prediction under uncertainty. Groundwater 53 (2), 305–316.

Cutler, A., Cutler, D.R., Stevens, J.R., 2012. Random forests. In: Ensemble machine learning. Springer, Boston, MA, pp. 157–175.

Dakou, E., D'heygere, T., Dedecker, A. P., Goethals, P. L., Lazaridou-Dimitriadou, M., & De Pauw, N, 2007. Decision tree models for prediction of macroinvertebrate taxa in the river Axios (Northern Greece). Aquat. Ecol. 41 (3), 399–411.

Darbandsari, P., Coulibaly, P., 2020. Introducing entropy-based Bayesian model averaging for streamflow forecast. J. Hydrol. 591, 125577.

Dietterich, T.G., 2002. Ensemble learning. The handbook of brain theory and neural networks 2, 110–125.

Diks, C.G., Vrugt, J.A., 2010. Comparison of point forecast accuracy of model averaging methods in hydrologic applications. Stoch. Env. Res. Risk Assess. 24 (6), 809–820.

Dong, Leihua, Xiong, Lihua, Yu, Kun-xia, 2013. Uncertainty Analysis of Multiple Hydrologic Models Using the Bayesian Model Averaging Method. Journal of Applied Mathematics 30, 701–710. https://doi.org/10.1155/2013/346045.

Duan, Q., Ajami, N.K., Gao, X., Sorooshian, S., 2007. Multi-model ensemble hydrologic prediction using Bayesian model averaging. Adv. Water Resour. 30 (5), 1371–1386.

Elkiran, G., Nourani, V., Abba, S.I., 2019. Multi-step ahead modelling of river water quality parameters using ensemble artificial intelligence-based approach. J. Hydrol. 577, 123962.

Erdal, H.I., Karakurt, O., 2013. Advancing monthly streamflow prediction accuracy of CART models using ensemble learning paradigms. J. Hydrol. 477, 119–128.

Francke, T., López-Tarazón, J.A., Schröder, B., 2008. Estimation of suspended sediment concentration and yield using linear models, random forests and quantile regression forests. Hydrological Processes: An International Journal 22 (25), 4892–4904.

Freund, Yoav, Schapire, Robert E., 1996. Experiments with a new boosting algorithm. Proc. 13th Int. Conf. Machine Learning (ICML) 148–156.

Freund, Y., Schapire, R.E., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. J. Comput. Syst. Sci. 55 (1), 119–139.

Friedel, M.J., Wilson, S.R., Close, M.E., Buscema, M., Abraham, P., Banasiak, L., 2020. Comparison of four learning-based methods for predicting groundwater redox status. J. Hydrol. 580, 124200.

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 1189–1232.

Friedman, J.H., 2002. Stochastic gradient boosting. Comput. Stat. Data Anal. 38 (4), 367–378.

Galelli, S., Castelletti, A., 2013. Assessing the predictive capability of randomized tree-based ensembles in streamflow modelling. Hydrol. Earth Syst. Sci. 17 (7), 2669–2684.

Georganos, S., Grippa, T., Vanhuysse, S., Lennert, M., Shimoni, M., Wolff, E., 2018. Very high resolution object-based land use–land cover urban classification using extreme gradient boosting. IEEE Geosci. Remote Sens. Lett. 15 (4), 607–611.

Geurts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees. Machine learning 63 (1), 3–42.

Ghorbani, M.A., Khatibi, R., Karimi, V., Yaseen, Z.M., Zounemat-Kermani, M., 2018. Learning from multiple models using artificial intelligence to improve model prediction accuracies: application to river flows. Water Resour. Manage. 32 (13), 4201–4215.

Govindaraju, R.S., 2000. Artificial neural networks in hydrology. II: hydrologic applications. J. Hydrol. Eng. 5 (2), 124–137.

He, X., Luo, J., Li, P., Zuo, G., Xie, J., 2020. A hybrid model based on variational mode decomposition and gradient boosting regression tree for monthly runoff forecasting. Water Resour. Manage. 34 (2), 865–884.

Heddam, S., 2021a. Intelligent Data Analytics Approaches for Predicting Dissolved Oxygen Concentration in River: Extremely Randomized Tree Versus Random Forest, MLPNN and MLR. In *Intelligent Data Analytics for Decision-Support Systems in Hazard Mitigation*. Springer, Singapore, pp. 89–107.

Heddam, S., 2021b. New formulation for predicting soil moisture content using only soil temperature as predictor: multivariate adaptive regression splines versus random forest, multilayer perceptron neural network, M5Tree, and multiple linear regression. In: Water Engineering Modeling and Mathematic Tools. Elsevier, pp. 45–62.

Heddam, S., Ptak, M., Zhu, S., 2020. Modelling of daily lake surface water temperature from air temperature: Extremely randomized trees (ERT) versus Air2Water, MARS, M5Tree, RF and MLPNN. J. Hydrol. 588, 125130.

Ho, T.K., 1998. The random subspace method for constructing decision forests. IEEE Trans. Pattern Anal. Mach. Intell. 20 (8), 832–844.

Hosseini, F.S., Choubin, B., Mosavi, A., Nabipour, N., Shamshirband, S., Darabi, H., Haghighi, A.T., 2020. Flash-flood hazard assessment using ensembles and Bayesian-based machine learning models: application of the simulated annealing feature selection method. Sci. Total Environ. 711, 135161.

Huang, H., Liang, Z., Li, B., Wang, D., Hu, Y., Li, Y., 2019. Combination of multiple data-driven models for long-term monthly runoff predictions based on Bayesian model averaging. Water Resour. Manage. 33 (9), 3321–3338.

Islam, A.R.M.T., Talukdar, S., Mahato, S., Kundu, S., Eibek, K.U., Pham, Q.B., Linh, N.T. T., 2020. Flood susceptibility modelling using advanced ensemble machine learning models. Geosci. Front. 12 (3), 101075.

Jeong, D.I., Kim, Y.O., 2005. Rainfall-runoff models using artificial neural networks for ensemble streamflow prediction. Hydrological Processes: An International Journal 19 (19), 3819–3835.

Kan, G., Yao, C., Li, Q., Li, Z., Yu, Z., Liu, Z., Liang, K., 2015. Improving event-based rainfall-runoff simulation using an ensemble artificial neural network based hybrid data-driven model. Stoch. Env. Res. Risk Assess. 29 (5), 1345–1370.

Kim, D., Yu, H., Lee, H., Beighley, E., Durand, M., Alsdorf, D.E., Hwang, E., 2019. Ensemble learning regression for estimating river discharges using satellite altimetry data: Central Congo River as a Test-bed. Remote Sens. Environ. 221, 741–755.

Kim, Sungwon, Alizamir, Meysam, Kim, Nam Won, Ozgur, Kisi, 2020. Bayesian Model Averaging: A Unique Model Enhancing Forecasting Accuracy for Daily Streamflow Based on Different Antecedent Time Series. Sustainability 12 (22), 9720. https://doi.org/10.3390/su12229720.

Kim, S.E., Seo, I.W., 2015. Artificial neural network ensemble modeling with exploratory factor analysis for streamflow forecasting. J. Hydroinf. 17 (4), 614–639.

Kisi, O., Heddam, S., 2019. Evaporation modelling by heuristic regression approaches using only temperature data. Hydrol. Sci. J. 64 (6), 653–672.

Kisi, O., Alizamir, M., Gorgij, A.D., 2020. Dissolved oxygen prediction using a new ensemble method. Environ. Sci. Pollut. Res. 1–15.

Kisi, O., Gorgij, A.D., Zounemat-Kermani, M., Mahdavi-Meymand, A., Kim, S., 2019. Drought forecasting using novel heuristic methods in a semi-arid environment. J. Hydrol. 578, 124053.

Lange, H., Sippel, S., 2020. Machine learning applications in hydrology. In: Forest-Water Interactions. Springer, Cham, pp. 233–257.

Laucelli, D., Giustolisi, O., Babovic, V., Keijzer, M., 2007. Ensemble modeling approach for rainfall/groundwater balancing. J. Hydroinf. 9 (2), 95–106.

Lee, S., Kim, J.C., Jung, H.S., Lee, M.J., Lee, S., 2017. Spatial prediction of flood susceptibility using random-forest and boosted-tree models in Seoul metropolitan city, Korea. Geomatics, Natural Hazards and Risk 8 (2), 1185–1203.

Li, H., Wen, G., Yu, Z., Zhou, T., 2013. Random subspace evidence classifier. Neurocomputing 110, 62–69.

Li, P.H., Kwon, H.H., Sun, L., Lall, U., Kao, J.J., 2010. A modified support vector machine based prediction model on streamflow at the Shihmen Reservoir. Taiwan. *International Journal of Climatology* 30 (8), 1256–1268.

Li, S., Ma, K., Jin, Z., Zhu, Y., 2016. In: July). A new flood forecasting model based on SVM and boosting learning algorithms. IEEE, pp. 1343–1348.

Li, X., Tsai, F.T.C., 2009. Bayesian model averaging for groundwater head prediction and uncertainty analysis using multimodel and multimethod. Water Resour. Res. 45 (9).

Li, Y., Liang, Z., Hu, Y., Li, B., Xu, B., Wang, D., 2020a. A multi-model integration method for monthly streamflow prediction: modified stacking ensemble strategy. J. Hydroinf. 22 (2), 310–326.

Li, Z., Chen, T., Wu, Q., Xia, G., Chi, D., 2020b. Application of penalized linear regression and ensemble methods for drought forecasting in Northeast China. Meteorol. Atmos. Phys. 132 (1), 113–130.

Liang, Z., Wang, D., Guo, Y., Zhang, Y., Dai, R., 2013. Application of Bayesian model averaging approach to multimodel ensemble hydrologic forecasting. J. Hydrol. Eng. 18 (11), 1426–1436.

Liu, S., Xu, J., Zhao, J., Xie, X., Zhang, W., 2014. Efficiency enhancement of a process-based rainfall–runoff model using a new modified AdaBoost. RT technique. *Applied Soft Computing* 23, 521–529.

Liu, Z., Merwade, V., 2018. Accounting for model structure, parameter and input forcing uncertainty in flood inundation modeling using Bayesian model averaging. J. Hydrol. 565, 138–149.

Lu, H., Ma, X., 2020. Hybrid decision tree-based machine learning models for short-term water quality prediction. Chemosphere 249, 126169.

Mei, Y., Hong, F., Jia, Z., Kang, Z., 2018. In: June). Debris Flow Forecasting of Northwest of Yunnan Province Based on LR, SVM, and RF Statistical Models. IEEE, pp. 1–7.

Melesse, A.M., Khosravi, K., Tiefenbacher, J.P., Heddam, S., Kim, S., Mosavi, A., Pham, B. T., 2020. River water salinity prediction using hybrid machine learning models. Water 12 (10), 2951.

Mert, A., Kılıç, N., Bilgili, E., 2016. Random subspace method with class separability weighting. Expert Systems 33 (3), 275–285.

Mirzaei, S., Vafakhah, M., Pradhan, B., Alavi, S.J., 2021. Flood susceptibility assessment using extreme gradient boosting (EGB). Iran. *Earth Science Informatics* 14 (1), 51–67.

Moeini, M., Shojaeizadeh, A., Geza, M., 2021. Supervised Machine Learning for Estimation of Total Suspended Solids in Urban Watersheds. Water 13 (2), 147.

Mosavi, A., Golshan, M., Janizadeh, S., Choubin, B., Melesse, A.M., Dineva, A.A., 2020. Ensemble models of GLM, FDA, MARS, and RF for flood and erosion susceptibility mapping: a priority assessment of sub-basins. Geocarto International 1–20.

Naghibi, S.A., Dolatkordestani, M., Rezaei, A., Amouzegari, P., Heravi, M.T., Kalantar, B., Pradhan, B., 2019. Application of rotation forest with decision trees as base classifier and a novel ensemble model in spatial modeling of groundwater potential. Environ. Monit. Assess. 191 (4), 248.

Naghibi, S.A., Hashemi, H., Berndtsson, R., Lee, S., 2020. Application of extreme gradient boosting and parallel random forest algorithms for assessing groundwater spring potential using DEM-derived factors. J. Hydrol. 589, 125197.

Najafi, M.R., Moradkhani, H., 2016. Ensemble combination of seasonal streamflow forecasts. J. Hydrol. Eng. 21 (1), 04015043.

Napiorkowski, M. J., Piotrowski, A. P., & Napiorkowski, J. J. (2014). Stream temperature forecasting by means of ensemble of neural networks: Importance of input variables and ensemble size. In River Flow (pp. 2017-2025).

Nguyen, P.T., Ha, D.H., Nguyen, H.D., Van Phong, T., Trinh, P.T., Al-Ansari, N., Prakash, I., 2020. Improvement of Credal Decision Trees Using Ensemble Frameworks for Groundwater Potential Modeling. Sustainability 12 (7), 2622.

Nhu, V.H., Khosravi, K., Cooper, J.R., Karimi, M., Kisi, O., Pham, B.T., Lyu, Z., 2020a. Monthly suspended sediment load prediction using artificial intelligence: testing of a new random subspace method. Hydrol. Sci. J. 1–12.

Nhu, V.H., Rahmati, O., Falah, F., Shojaei, S., Al-Ansari, N., Shahabi, H., Ahmad, B.B., 2020b. Mapping of Groundwater Spring Potential in Karst Aquifer System Using Novel Ensemble Bivariate and Multivariate Models. Water 12 (4), 985.

Ni, L., Wang, D., Wu, J., Wang, Y., Tao, Y., Zhang, J., Liu, J., 2020. Streamflow forecasting using extreme gradient boosting model coupled with Gaussian mixture model. J. Hydrol. 586, 124901.

Nolan, B.T., Fienen, M.N., Lorenz, D.L., 2015. A statistical learning framework for groundwater nitrate models of the Central Valley, California, USA. J. Hydrol. 531, 902–911.

Oehler, F., Coco, G., Green, M.O., Bryan, K.R., 2012. A data-driven approach to predict suspended-sediment reference concentration under non-breaking waves. Cont. Shelf Res. 46, 96–106.

Pai, P.F., Li, L.L., Hung, W.Z., Lin, K.P., 2014. Using ADABOOST and rough set theory for predicting debris flow disaster. Water Resour. Manage. 28 (4), 1143–1155.

Pham, B.T., Jaafari, A., Van Phong, T., Yen, H.P.H., Tuyen, T.T., Van Luong, V., Foong, L. K., 2021. Improved flood susceptibility mapping using a best first decision tree integrated with ensemble learning techniques. Geosci. Front. 12 (3), 101105.

Phukoetphim, P., & Shamseldin, A. Y. (2013). The use of stochastic gradient boosting method for multi-model combination of rainfall-runoff models. In International Conference on Innovations in Engineering and Technology (ICIET '2013) (pp. 275-280).

Polikar, R., 2012. Ensemble learning. In: Ensemble machine learning. Springer, Boston, MA, pp. 1–34.

Qu, B., Zhang, X., Pappenberger, F., Zhang, T., Fang, Y., 2017. Multi-model grand ensemble hydrologic forecasting in the Fu river basin using Bayesian model averaging. Water 9 (2), 74.

Rajaee, T., Khani, S., Ravansalar, M., 2020. Artificial intelligence-based single and hybrid models for prediction of water quality in rivers: A review. Chemometrics and Intelligent Laboratory Systems 200, 103978.

Rizeei, H.M., Pradhan, B., Saharkhiz, M.A., Lee, S., 2019. Groundwater aquifer potential modeling using an ensemble multi-adoptive boosting logistic regression technique. J. Hydrol. 579, 124172.
Rodriguez, J.J., Kuncheva, L.I., Alonso, C.J., 2006. Rotation forest: A new classifier ensemble method. IEEE Trans. Pattern Anal. Mach. Intell. 28 (10), 1619–1630.
Rodriguez-Galiano, V., Mendes, M.P., Garcia-Soldado, M.J., Chica-Olmo, M., Ribeiro, L., 2014. Predictive modeling of groundwater nitrate pollution using Random Forest and multisource variables related to intrinsic and specific vulnerability: A case study in an agricultural setting (Southern Spain). Sci. Total Environ. 476, 189–206.
Rojas, R., Feyen, L., Dassargues, A., 2008. Conceptual model uncertainty in groundwater modeling: Combining generalized likelihood uncertainty estimation and Bayesian model averaging. Water Resour. Res. 44 (12).
Sahour, H., Gholami, V., Vazifedan, M., 2020. A comparative analysis of statistical and machine learning techniques for mapping the spatial distribution of groundwater salinity in a coastal aquifer. J. Hydrol. 591, 125321.
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. J. Hydrol. 590, 125241.
Schapire, R.E., 2003. The boosting approach to machine learning: An overview. In: Nonlinear estimation and classification. Springer, New York, NY, pp. 149–171.
Schick, S., Rössler, O.K., Weingartner, R., 2018. Monthly streamflow forecasting at varying spatial scales in the Rhine basin. Hydrol. Earth Syst. Sci. 22 (2), 929–942.
Shadkani, S., Abbaspour, A., Samadianfard, S., Hashemi, S., Mosavi, A., Band, S.S., 2020. Comparative study of multilayer perceptron-stochastic gradient descent and gradient boosted trees for predicting daily suspended sediment load: The case study of the Mississippi River, US. Int. J. Sedim. Res.
Shafizadeh-Moghadam, H., Valavi, R., Shahabi, H., Chapi, K., Shirzadi, A., 2018. Novel forecasting approaches using combination of machine learning and statistical models for flood susceptibility mapping. J. Environ. Manage. 217, 1–11.
Shamaei, E., Kaedi, M., 2016. Suspended sediment concentration estimation by stacking the genetic programming and neuro-fuzzy predictions. Appl. Soft Comput. 45, 187–196.
Sharafati, A., Asadollah, S.B.H.S., Hosseinzadeh, M., 2020a. The potential of new ensemble machine learning models for effluent quality parameters prediction and related uncertainty. Process Saf. Environ. Prot. 140, 68–78.
Sharafati, A., Asadollah, S.B.H.S., Neshat, A., 2020b. A new artificial intelligence strategy for predicting the groundwater level over the Rafsanjan aquifer in Iran. J. Hydrol. 591, 125468.
Sharafati, A., Asadollah, Haji Seyed, S. B., Motta, D., & Yaseen, Z. M, 2020c. Application of newly developed ensemble machine learning models for daily suspended sediment load prediction and related uncertainty analysis. Hydrol. Sci. J. 1–21.
Sharghi, E., Nourani, V., Behfar, N., 2018. Earthfill dam seepage analysis using ensemble artificial intelligence based modeling. J. Hydroinf. 20 (5), 1071–1084.
Sharma, P., Machiwal, D., & Jha, M. K. (2019). Overview, Current Status, and Future Prospect of Stochastic Time Series Modeling in Subsurface Hydrology.
Sharma, S.K., Tiwari, K.N., 2009. Bootstrap based artificial neural network (BANN) analysis for hierarchical prediction of monthly runoff in Upper Damodar Valley Catchment. J. Hydrol. 374 (3–4), 209–222.
Shin, J., Yoon, S., Cha, Y., 2017. Prediction of cyanobacteria blooms in the lower Han River (South Korea) using ensemble learning algorithms. DESALINATION AND WATER TREATMENT 84, 31–39.
Shu, C., Burn, D.H., 2004. Artificial neural network ensembles and their application in pooled flood frequency analysis. Water Resour. Res. 40 (9).
Sihag, Parveen, Al-Janabi, Ahmed Mohammed Sami, Alomari, Nashwan K., 2021. Aminuddin Ab Ghani, and Somvir Singh Nain. "Evaluation of tree regression analysis for estimation of river basin discharge.". Modeling Earth Systems and Environment 1–13.
Singh, G., Panda, R.K., 2015. Bootstrap-based artificial neural network analysis for estimation of daily sediment yield from a small agricultural watershed. International Journal of Hydrology Science and Technology 5 (4), 333–348.
Singh, K.P., Gupta, S., Mohan, D., 2014. Evaluating influences of seasonal variations and anthropogenic activities on alluvial groundwater hydrochemistry using ensemble learning approaches. J. Hydrol. 511, 254–266.
Sun, W., Li, Z., 2020. Hourly PM2. 5 concentration forecasting based on feature extraction and stacking-driven ensemble model for the winter of the Beijing-Tianjin-Hebei area. Atmospheric. Pollution Research.
Sun, W., Trevor, B., 2018. A stacking ensemble learning framework for annual river ice breakup dates. J. Hydrol. 561, 636–650.
Sun, W., Lv, Y., Li, G., Chen, Y., 2020. Modeling River Ice Breakup Dates by k-Nearest Neighbor Ensemble. Water 12 (1), 220.
Ting, K.M., Witten, I.H., 1997. Stacking bagged and dagged models. In: In Proceedings of the 14th international conference on machine learning, pp. 367–375.
Tiwari, M.K., Chatterjee, C., 2010. Development of an accurate and reliable hourly flood forecasting model using wavelet–bootstrap–ANN (WBANN) hybrid approach. J. Hydrol. 394 (3–4), 458–470.
Tiwari, M.K., Chatterjee, C., 2011. A new wavelet–bootstrap–ANN hybrid model for daily discharge forecasting. J. Hydroinf. 13 (3), 500–519.
Turco, M., Ceglar, A., Prodhomme, C., Soret, A., Toreti, A., Francisco, J.D.R., 2017. Summer drought predictability over Europe: empirical versus dynamical forecasts. Environ. Res. Lett. 12 (8), 084006.
Tyralis, H., Papacharalampous, G., & Langousis, A. (2019). Super learning for daily streamflow forecasting: Large-scale demonstration and comparison with multiple machine learning algorithms. arXiv preprint arXiv:1909.04131.
Tyralis, H., Papacharalampous, G., Langousis, A., 2020. Super ensemble learning for daily streamflow forecasting: large-scale demonstration and comparison with multiple machine learning algorithms. Neural Comput. Appl. 1–16.
Venkatesan, E., Mahindrakar, A.B., 2019. Forecasting floods using extreme gradient boosting–a new approach. International Journal of Civil Engineering and Technology 10 (2), 1336–1346.
Wang, X., Yen, H., Jeong, J., Williams, J.R., 2015. Accounting for conceptual soil erosion and sediment yield modeling uncertainty in the APEX model using Bayesian model averaging. J. Hydrol. Eng. 20 (6), C4014010.
Wang, Y., Hong, H., Chen, W., Li, S., Panahi, M., Khosravi, K., Costache, R., 2019. Flood susceptibility mapping in Dingnan County (China) using adaptive neuro-fuzzy inference system with biogeography based optimization and imperialistic competitive algorithm. J. Environ. Manage. 247, 712–729.
Wolpert, D.H., 1992. Stacked generalization. *Neural networks* 5 (2), 241–259.
Wright, J.H., 2008. Bayesian model averaging and exchange rate forecasts. Journal of Econometrics 146 (2), 329–341.
Wu, Y., Ke, Y., Chen, Z., Liang, S., Zhao, H., Hong, H., 2020. Application of alternating decision tree with AdaBoost and bagging ensembles for landslide susceptibility mapping. Catena 187, 104396.
Xu, J., Zhang, Q., Liu, S., Zhang, S., Jin, S., Li, D., ... & Li, H., 2020. Ensemble learning of daily river discharge modeling for two watersheds with different climates. Atmos. Sci. Lett. e1000.
Xu, L., Chen, N., Zhang, X., 2018. A comparison of large-scale climate signals and the North American Multi-Model Ensemble (NMME) for drought prediction in China. J. Hydrol. 557, 378–390.
Yariyan, P., Janizadeh, S., Van Phong, T., Nguyen, H.D., Costache, R., Van Le, H., Tiefenbacher, J.P., 2020. Improvement of best first decision trees using bagging and dagging ensembles for flood probability mapping. Water Resour. Manage. 34 (9), 3037–3053.
Yin, J., Medellín-Azuara, J., Escriva-Bou, A., Liu, Z., 2021. Bayesian machine learning ensemble approach to quantify model uncertainty in predicting groundwater storage change. Sci. Total Environ. 144715.
Yu, X., Wang, Y., Wu, L., Chen, G., Wang, L., Qin, H., 2020. Comparison of support vector regression and extreme gradient boosting for decomposition-based data-driven 10-day streamflow forecasting. J. Hydrol. 582, 124293.
Yuan, W., Liu, M., Wan, F., 2019. Calculation of critical rainfall for small-watershed flash floods based on the HEC-HMS hydrological model. Water Resour. Manage. 33 (7), 2555–2575.
Zhai, J., Mondal, S.K., Fischer, T., Wang, Y., Su, B., Huang, J., Uddin, M.J., 2020. Future drought characteristics through a multi-model ensemble from CMIP6 over South Asia. Atmos. Res. 246, 105111.
Zhang, C., Ma, Y. (Eds.), 2012. Ensemble machine learning: methods and applications. Springer Science & Business Media.
Zhou, J., Peng, T., Zhang, C., Sun, N., 2018. Data pre-analysis and ensemble of various artificial neural networks for monthly streamflow forecasting. Water 10 (5), 628.
Zhou, Z.H., 2009. Ensemble Learning. Encyclopedia of biometrics 1, 270–273.
Zhu, S., Nyarko, E.K., Hadzima-Nyarko, M., Heddam, S., Wu, S., 2019. Assessing the performance of a suite of machine learning models for daily river water temperature prediction. PeerJ 7, e7065.
Zounemat-Kermani, M., 2017. Assessment of several nonlinear methods in forecasting suspended sediment concentration in streams. Hydrol. Res. 48 (5), 1240–1252.
Zounemat-Kermani, M., Scholz, M., 2013. Computing air demand using the Takagi-Sugeno model for dam outlets. Water 5 (3), 1441–1456.
Zounemat-Kermani, M., Kisi, O., Piri, J., Mahdavi-Meymand, A., 2019. Assessment of Artificial Intelligence-Based Models and Metaheuristic Algorithms in Modeling Evaporation. J. Hydrol. Eng. 24 (10), 04019033.
Zounemat-Kermani, M., Matta, E., Cominola, A., Xia, X., Zhang, Q., Liang, Q., Hinkelmann, R., 2020a. Neurocomputing in Surface Water Hydrology and Hydraulics: A Review of Two Decades Retrospective, Current Status and Future Prospects. J. Hydrol. 125085.
Zounemat-Kermani, M., Stephan, D., Barjenbruch, M., Hinkelmann, R., 2020b. Ensemble data mining modeling in corrosion of concrete sewer: A comparative study of network-based (MLPNN & RBFNN) and tree-based (RF, CHAID, & CART) models. Adv. Eng. Inf. 43, 101030.