IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 1, 2021, accepted August 31, 2021, date of publication September 9, 2021, date of current version September 17, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3111287

Development of an Extreme Gradient Boosting Model Integrated With Evolutionary Algorithms for Hourly Water Level Prediction

DUC HAI NGUYEN[1,2], XUAN HIEN LE[2], JAE-YEONG HEO[1], AND DEG-HYO BAE[1]
[1]Department of Civil and Environmental Engineering, Sejong University, Gwangjin-Gu, Seoul 143-747, South Korea
[2]Faculty of Water Resources Engineering, Thuyloi University, Dong Da, Hanoi 116705, Vietnam

Corresponding author: Deg-Hyo Bae (dhbae@sejong.ac.kr)

This work was supported by the Faculty Research Fund of Sejong University in 2021.

ABSTRACT The establishment of reliable water level prediction models is vital for urban flood control and planning. In this paper, we develop hybrid models (GA-XGBoost and DE-XGBoost) that couple two evolutionary models, a genetic algorithm (GA) and a differential evolution (DE) algorithm, with the extreme gradient boosting (XGBoost) model for hourly water level prediction. The Jungrang urban basin located on the Han River, South Korea, was selected as a case study for the proposed models. Hourly rainfall and water level data were collected between 2003 and 2020 to construct and evaluate the performance of the selected models. To compare the prediction efficiency, two other tree-based models were chosen: classification and registration tree (CART) and random forest (RF) models. A comparison of the results showed that two hybrid models, GA-XGBoost and DE-XGBoost, outperformed RF and CART in the multistep-ahead prediction of water level, and the relative errors of the hybrid model ranged from [2.18%-9.21%], compared to [3.76%-10.41%] and [2.99%-11.88%] for the RF and CART, respectively. Reliable performance was also supported by other measures. In general, the GA-XGBoost and DE-XGBoost models displayed relatively similar performance despite their small differences. The CART model was not preferable for multistep-ahead water level predictions, even though it yielded the lowest Akaike information criterion (AIC) value. This study verifies that despite having some drawbacks when considering long step-ahead prediction and model complexity, hybrid XGBoost models might be superior to many existing models for hourly water level prediction.

INDEX TERMS Extreme gradient boosting, evolutionary algorithms, water level prediction, tree-based model, urban floods.

I. INTRODUCTION

Floods are among the greatest risks in most cities around the world. Due to hydrometeorological and hydrological variability induced by climate change, urban floods are more complex now than in previous decades [1], [2]. Therefore, flood management is becoming a critical challenge, especially in developed cities. Reliable urban flood prediction for heavy rainfall events is vital for alleviating the damage to urban basins [3]. Consequently, high-accuracy water level predictions for a fine time step are necessary for urban areas and are being given significant attention from scientists.

To achieve accurate predictions, various approaches have been established and applied. These attempts can be divided into two groups. The first group involves the coupling of hydrological and meteorological forecasting models for predictions based on physical rainfall-runoff formulations [4]–[6]. These methods generally use simplified assumptions for hydrological processes and require forecasted hydrometeorological data. The second group includes data-driven methods, such as statistical approaches and machine learning approaches, that do not require excessive data for hydrological processes and are not difficult to apply.

The associate editor coordinating the review of this manuscript and approving it for publication was Xujie Li.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

Data-driven methods mainly use the relevant features of past data to make hydrological predictions.

Machine learning models have been widely used in many areas of hydrology in recent years due to their superior capabilities in learning the details of complex hydrological processes. Various machine learning algorithms have been implemented and developed, including the adaptive neuro-fuzzy inference system (ANFIS) [7], [8], support vector machine (SVM) [9], [10], neural networks-based model (i.e., recurrent neural network (RNN) [11] and long short-term memory (LSTM) [12]–[16]), auto regressive moving average (ARMA) [17], [18] and genetic programming [19]–[22]. Machine learning algorithms have provided reliable performances in terms of forecasting water levels and streamflow and simulating rainfall-runoff mechanisms.

Unlike the abovementioned machine learning models, tree-based models are computationally cheap[23]. In addition, there is no requirement regarding the distribution of predictors in tree-based models. Additionally, whereas a neural network is almost identical to a black box that provides results without a clear explanation, tree-based models provide an understandable interpretation for visualization [24]. Classification and regression trees (CARTs) have been used to solve hydrological problems, such as runoff generation [25] and streamflow forecasting [26]. Random forest (RF) algorithms have also been used for runoff and streamflow simulation [27]–[29]. Extreme gradient boosting (XGBoost) was first apprised by Chen and Guestrin [30] and has received considerable attention from scientists for machine learning applications. Whereas an RF is an ensemble tree-based algorithm, XGBoost is based on boosting trees that use a gradient descent algorithm. The XGBoost model uses additive training strategies to consider all the outputs of weak learners to create a strong learner. XGBoost leverages strengths from two algorithms: gradient boosting (GB) and the decision tree (DT) algorithm. Therefore, this scalable algorithm might provide good results and visualizations in terms of effective water level predictions in urban areas. Recently, a few works have explored the performance of XGBoost in the hydrological domain and obtained promising results. Ni *et al.* [23] used the XGBoost method integrated with the Gaussian mixture approach for forecasting monthly streamflow; they concluded that the advanced tree-based model outperformed a support vector machine (SVM) and suggested that the new model be applied in water management as a superior option because of its reliable performance Hadi *et al.* [31] integrated XGBoost and an extreme learning machine (ELM) to model monthly streamflow. These previous studies successfully applied models for monthly streamflow predictions in large river basins. However, the capabilities of tree-based models at the urban basin scale and at hourly time steps are still poorly understood. Therefore, we investigated the performance of the XGBoost model in hourly water level prediction for a small urban basin in this study.

The development of the XGBoost model requires the internal optimization of hyperparameters. Hybrid models can provide a solution in this type of optimization process [32]. The present study integrates the XGBoost model with two evolutionary algorithms: a genetic algorithm (GA) and a differential evolution (DE) algorithm. The GA and DE algorithms were used to optimize the hyperparameters during the training stage and then generate an optimal set of parameters for the testing stage. The GA is a type of stochastic search algorithm that is based on the fundamental concepts of evolution and natural selection [33]. The GA mimics the biological process of evolution through selection, crossover, and mutation [34]. This algorithm has received significant attention in recent studies [19], [35]–[38]. The DE algorithm was developed based on the GA and is a well-designed metaheuristic method for the global optimization of non-continuous or nondifferentiable functions. The DE algorithm is suitable for optimizing continuous variables in multiple dimensions. Some hydrology-related studies have recently applied the DE algorithm in streamflow simulation [39], flood assessment [40], hydraulic design [41], [42] and water capacity modeling [43], [44].

The objective of this study is to investigate the performance of the hybrid XGBoost models GA-XGBoost and DE-XGBoost for predicting hourly multistep-ahead water levels in an urban basin. In addition, the performance of the hybrid XGBoost models is compared to that of two tree-based models, including a CART and an RF. Detailed information about the study area and the data used are given in section 2. The different algorithms used in this work are provided in section 3. The model configuration and implementation schemes are presented in section 4. The fifth section illustrates the performance of the algorithms, and a discussion on the results is provided. Finally, the conclusions are discussed at the end of this article.

II. STUDY AREA AND DATA PROCESSING

A. STUDY AREA

The Jungrang basin is located between 37°32' and 37°39' latitude and from 126°58' to 127°9' longitude in the lower part of the Han River basin. The Jungrang basin includes an upstream part and a downstream part, which jointly cover an area of 299.87 km^2. The downstream part of the Jungrang basin is situated in the Seoul metropolitan area, which is one of the largest cities in South Korea. In cities, urban flooding typically occurs within a few hours of a strong precipitation event. Therefore, this basin is vulnerable to floods and flood destruction following heavy rain events, such as the historical precipitation events in 2006, 2010, and 2011. The climate in this region is a continental monsoon climate, which is characterized by an uneven distribution of rainfall throughout the year. The highest incidence of flood events is from June to September annually. The locations of the Jungrang basin on the Korean Peninsula, rainfall stations, and water level stations are presented in Fig. 1; ten ground rainfall stations and three water level stations are denoted by red triangles and black pentagons (Fig. 1(b)), respectively.

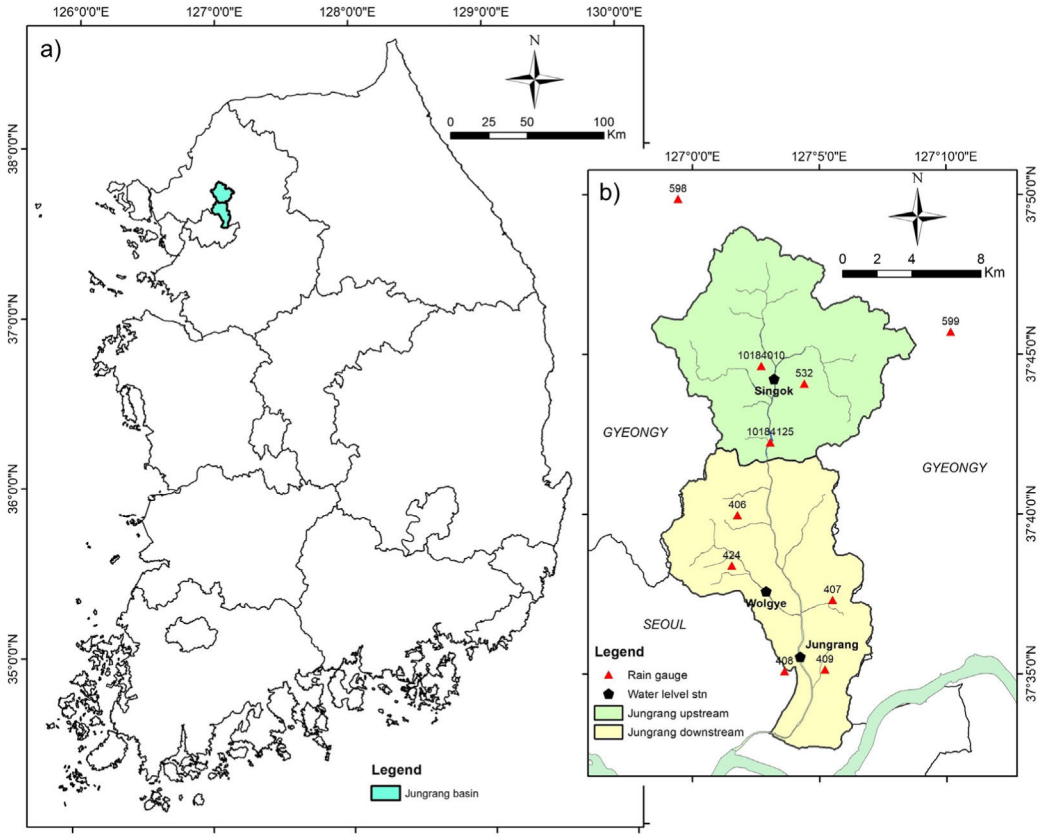

FIGURE 1. (a) Location of the Jungrang Basin in South Korea. (b) Jungrang Basin and locations of the rain gauges and water level stations.

B. DATA PROCESSING

Measured rainfall data from 10 selected ground gauges with automatic weather systems were collected by the Korean Meteorological Administration (KMA) and Water Resources Management Information System (WAMIS). The KMA stations include IDs 406, 407, 408, 409, 424, 532, 598, and 599. The WAMIS stations include IDs 10184010 and 10184125. All the selected rain gauges record information with a temporal resolution of 1 hour. Hourly observed water level data were collected by the Han River Flood Control Office (HRFCO) at the Jungrang, Wolgye, and Singok stations. Series of rainfall and water level data were collected and processed from 2003 to 2020. The heavy rainfall events that normally occur from June to September were selected. Table 1 shows the selected events and the corresponding information used in this study. The presented minimum and maximum water level series (Table 1) were measured at Jungrang station. The events numbered from 1 to 36 in Table 1 occurred between 2003 and 2017 and were chosen for calibration and validation in the training stage. The period from 2018 to 2020, which included six heavy rainfall events (from 37 to 42 in Table 1), was chosen as the testing stage.

In this study, we used the mean areal precipitation (MAP) in the upstream and downstream subbasins as the main inputs of the tree-based models. A quality control process for rain gauge data was implemented to fill missing values and verify abnormal values by using the inverse distance weighting (IDW) approach [45]. This process makes the data completed and realistic, then the models can learn and perform reasonable predictions. The hourly observed MAP was calculated by using the Thiessen polygon technique. Table 2 presents the descriptive statistics and the configurations of inputs and outputs used in this study. There were nine input variables for the models at each time step (t), including the MAP in the upstream and downstream subbasins at time steps (t, $t-1$, and $t-2$) [39] and the water level at time step t at three stations. The output of the different time-step models is the water level at Jungrang station for the next time steps (e.g., $t+1$, $t+2$, $t+3$, $t+4$, $t+5$, and $t+6$).

III. METHODOLOGY

A. CLASSIFICATION AND REGRESSION TREES (CARTs)

First proposed by Breiman *et al.* [46], CARTs are common machine learning algorithms used for data classification and regression. CARTs do not use any assumptions regarding parameters but instead focus on the repetitive division of the dataset to build a decision tree. A CART can handle various inputs, such as numerical, categorical, and binary inputs; therefore, CARTS can be used for hydrological datasets [25]. The CART algorithm uses the binary recursive partitioning

TABLE 1. Summary of rainfall events used in this research.

No.	Stage	Year	Durations (month/day:hour)	Min. water level (m)	Max. water level (m)	Max. rainfall (mm/h)
1	Training & Validation	2003	05/06:00-05/15:23	0.56	2.22	15.5
2			07/21:12-08/01:23	0.68	3.55	28.8
3			08/18:00-09/04:23	0.62	3.75	46.9
4		2004	09/17:00-09/24:00	0.72	3.82	32.2
5			07/11:12-07/24:23	0.69	2.54	22.2
6			06/26:00-07/08:23	0.42	2.40	26.8
7		2005	07/27:12-08/07:12	0.53	2.33	43.1
8			08/10:10-08/17:00	0.70	2.35	16.1
9			09/13:00-09/16:12	0.53	1.99	28.5
10		2006	07/11:00-07/24:23	0.61	5.19	34.5
11			07/26:12-08/03:00	0.70	3.65	31.3
12		2007	07/10:00-07/15:00	0.49	2.22	47.1
13			08/06:12-08/14:00	0.70	2.42	15.7
14		2008	07/18:12-07/23:12	0.51	2.46	21.1
15			07/23:22-07/30:23	0.78	3.67	28.9
16		2009	07/08:12-07/24:00	0.51	4.65	35.9
17			08/11:00-08/19:00	0.52	3.79	33.2
18		2010	07/16:00-07/22:23	0.51	2.16	18.9
19			08/14:12-08/18:00	0.79	3.14	45.6
20			08/29:00-09/01:00	0.89	3.34	23.4
21		2011	09/09:00-09/16:00	0.77	3.65	49.4
22			06/28:12-07/06:12	0.73	3.61	31.5
23			07/24:12-08/07:00	0.65	5.21	37.6
24		2012	07/05:00-07/10:00	0.50	2.78	21.2
25			07/18:12-07/24:23	0.71	2.69	22.7
26			08/14:00-08/18:12	0.48	2.75	20.0
27			08/19:12-27/08:12	0.64	2.68	25.9
28		2013	09/16:12-09/21:00	0.64	2.61	15.6
29			07/07:12-07/20:00	0.54	2.98	29.4
30			07/21:12-07/27:00	0.76	2.32	22.2
31		2016	07/01:00-07/04:00	0.48	2.55	29.8
32			07/04:12-07/10:00	0.59	4.67	40.3
33		2017	07/01:12-07/06:00	0.45	3.06	30.0
34			07/09:00-07/14:00	0.68	3.06	23.9
35			07/22:12-07/27:00	0.62	2.74	38.9
36			08/19:12-08/30:00	0.66	2.58	30.8
37	Testing	2018	08/28:12-09/03:12	0.56	5.42	54.5
38		2019	07/25:12-07/29:23	0.46	2.02	15.8
39		2020	07/23:00-07/27:23	0.58	2.39	17.8
40			08/02:03-08/05:06	0.90	3.97	38.4
41			08/05:07-08/08:18	0.90	3.70	21.0
42			08/08:20-08/14:00	0.84	3.18	24.5

approach to split datasets until they become homogeneous clusters set based on a certain threshold. This algorithm can produce an output with the structure of a hierarchical binary model that is easy to visualize and understand.

CARTs use the Gini coefficient and decreasing variance concepts in the binary classification approach [46]. In this study, the variables in the datasets are sequential; therefore, the Gini variance was utilized to develop a regression tree. Detailed information regarding CART algorithms was provided by Breiman *et al.* [46], Han *et al.* [47], and Lee and Kim [25].

B. RANDOM FOREST (RF)

First proposed by Breiman [48], the random forest (RF) approach is an ensemble machine learning method involving decision trees. The RF algorithm provides many advantages over other algorithms, including avoiding overfitting, capturing nonlinearity, and using a small set of model parameters. In the regression method, the RF model is trained many times with bootstrap samples, and the outputs of every tree are averaged to obtain the expected value. In an RF decision tree, a subset of variables used for optimizing objective functions and splitting each node is arbitrarily chosen, and this subset is autonomous among trees with different nodes. The process in which multiple training datasets are generated by the bootstrap resampling of the initial training dataset is called bagging or bootstrap aggregation. Therefore, many trees are normally used, and the number of trees corresponds to the *ntree* parameter in the RF algorithm. The *mtry* parameter in the RF model is related to the splitting process at each node based on an arbitrary choice of a subset of variables. A RF regression model issues sets of parent and child nodes for each bootstrap sample until the stopping condition is met with reference to the minimum node size parameter. Detailed information regarding RF algorithms and parameters was provided by Breiman [48] and in recent studies [29], [49].

TABLE 2. The inputs and outputs of the models.

	Notation	Explanation	Unit	Mean	Std
Input	$MAPu_t$	Upstream MAP at time t	mm/h		
	$MAPu_{t-1}$	Upstream MAP at time t-1	mm/h	0.64	2.60
	$MAPu_{t-2}$	Upstream MAP at time t-2	mm/h		
	$MAPd_t$	Downstream MAP at time t	mm/h		
	$MAPd_{t-1}$	Downstream MAP at time t-1	mm/h	0.59	2.33
	$MAPd_{t-2}$	Downstream MAP at time t-2	mm/h		
	WLS_t	WL in Singok at time t	m	0.82	0.38
	WLW_t	WL in Wolgye at time t	m	0.58	0.42
	WLJ_t	WL in Jungrang at time t	m	0.89	0.43
Output	WLJ_{t+1}	WL in Jungrang at time t+1	m	0.89	0.43
	WLJ_{t+2}	WL in Jungrang at time t+2	m		
	WLJ_{t+3}	WL in Jungrang at time t+3	m		
	WLJ_{t+4}	WL in Jungrang at time t+4	m		
	WLJ_{t+5}	WL in Jungrang at time t+5	m		
	WLJ_{t+6}	WL in Jungrang at time t+6	m		

Note: WL denotes the water level at the stations. Std is the standard deviation of the variables.

C. EXTREME GRADIENT BOOSTING (XGBOOST)

First proposed by Chen and Guestrin [30], extreme gradient boosting (XGBoost) is an efficient and scalable machine learning method involving tree boosting systems based on the original gradient boosting framework of Friedman [50]. Compared to the original gradient boosting model, XGBoost can perform parallelization processes in constructed boost trees to independently generate branches. XGBoost uses a CART ensemble to fit samples of training data. Each CART is associated with an autonomous decision rule for a binary tree, and each leaf node yields a predictive score. The algorithm outputs the sum of the corresponding node values for a given input. Fig. 2 shows a simple example of the XGBoost algorithm. If the given input x is for an 18-year-old man who uses a personal computer each day, then the output of the mapping tree (shown in Fig. 2) is $\hat{y}(x) = 2 + 0.9 = 2.9$. In general, the predicted score can be written as follows:

$$\hat{y_k} = \sum_{i=1}^{n} g_i(x_k), g_i \in G \quad (1)$$

where $\hat{y_k}$ denotes the predicted score, g_i denotes a CART, x_k denotes the input vector, n denotes the number of CARTs, and G represents the CART space.

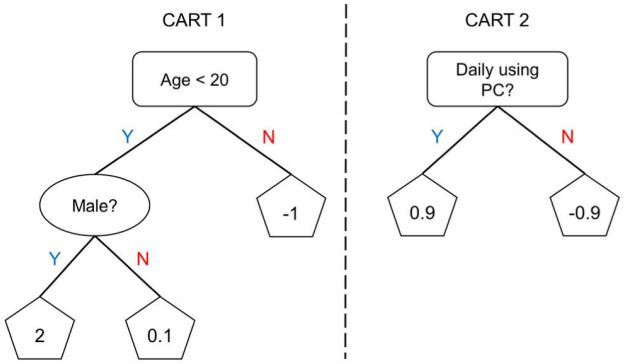

FIGURE 2. A classical instance of the XGBoost model.

A regularized objective function in XGBoost that includes two parts is optimized to define the CARTs as follows:

$$\Theta(\phi) = \sum_{k=1}^{K} L\left(y_k, \hat{y_k}\right) + \sum_{i=1}^{N} \Omega(g_i),$$
$$\Omega(g) = \gamma V + 0.5\lambda \|w\|^2 \quad (2)$$

where L presents the training loss function between the true value y and predicted value $\hat{y}$; Ω denotes regularization penalty scaling; V denotes the number of leaves in a CART; and w presents the score vectors for leaves. The regularization term is used to control the complexity of the model and avoid overfitting.

The XGBoost model is formally trained in an additive way. At the t-th iteration, the t-th CART is added to minimize the following objective:

$$\Theta^{(t)} = \sum_{k=1}^{K} L\left(y_k, \hat{y_k}^{(t-1)} + g_t(x_k)\right) + \Omega(g_t) \quad (3)$$

The model uses second-order Taylor expansion to simplify the objective:

$$g(x) = g(a) + g'(a)(x-a) + \frac{g''(a)}{2}(x-a)^2 \quad (4)$$

By applying Eq. (4) with x as the objective $\Theta^{(t)}$ (Eq. (3)) and $\hat{y_k}^{(t-1)}$ as a, the function objective is estimated as follows:

$$\Theta^{(t)} \cong \sum_{k=1}^{K} L\left(y_k, \hat{y_k}^{(t-1)} + f_k g_t(x_k) + \frac{s_k g_t^2(x_k)}{2}\right) + \Omega(g_t) \quad (5)$$

where f_k and s_k denote the first- and second-order formulations of the loss function, respectively. A simplified formulation of the objective can be obtained after removing the constant terms:

$$g\Theta^{(t)} \cong \sum_{k=1}^{K} \left(f_k g_t(x_k) + \frac{s_k g_t^2(x_k)}{2}\right) + \Omega(g_t) \quad (6)$$

Let $I_j = \{k \mid q(x_k) = j\}$ be the instance set for leaf j. Eq. (6) can be rewritten by expanding Ω in the following equation:

$$\Theta^{(t)} \cong \sum_{k=1}^{K} \left(f_k g_t(x_k) + \frac{s_k g_t^2(x_k)}{2}\right) + \gamma V + 0.5\lambda \sum_{j=1}^{V} w_j^2$$
$$\cong \sum_{j=1}^{V} \left[\left(\sum_{k\in I_j} f_k\right) w_j + \frac{1}{2}\left(\sum_{k\in I_j} s_k + \lambda\right) w_j^2\right] + \gamma V \quad (7)$$

By calculating the derivatives of Eq. (7) according to w_j and equating them to zero, the optimal weight w_j^* of leaf j is obtained as follows:

$$w_j^* = \frac{\sum_{k\in I_j} f_k}{\sum_{k\in I_j} s_k + \lambda} \quad (8)$$

The simplified objective can be rewritten as:

$$\Theta^{(t)} \cong -\frac{1}{2} \sum_{j=1}^{V} \frac{\left(\sum_{k\in I_j} f_k\right)^2}{\sum_{k\in I_j} s_k + \lambda} + \gamma V \quad (9)$$

Let I_R and I_L be the instance sets for right and left nodes after splitting. Given $I = I_R \cup I_L$, loss reduction after splitting is estimated by:

$$\Theta_{split} = \frac{1}{2}\left[\frac{\left(\sum_{k\in I_R} f_k\right)^2}{\sum_{k\in I_R} s_k + \lambda} + \frac{\left(\sum_{k\in I_L} f_k\right)^2}{\sum_{k\in I_L} s_k + \lambda} - \frac{\left(\sum_{k\in I} f_k\right)^2}{\sum_{k\in I} s_k + \lambda}\right] - \gamma \quad (10)$$

D. GENETIC ALGORITHM (GA)

The genetic algorithms (GA) was first proposed by Holland [51] and further expanded by Goldberg [52]. The GA approach is inspired by biological evolution phenomena, including natural selection, chromosomal crossover, and genetic mutation. GAs use stochastic optimization to explore the best values in a complex search space. This approach can be used to optimize continuous and discontinuous functions, whether constrained or unconstrained.

Fig. 3 shows a flowchart of the GA used in this study. The GA operators are described as follows.

i. *Selection:* An initial arbitrary population of a given size is generated at the beginning of the evolution process; at step $k = 0$, we obtain $\left\{\theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_n^{(0)}\right\}$. At step k, the fitness $f\left(\theta_i^k\right)$ of each element in the population is calculated, and probabilities p_i^k are assigned to all elements as follows:

$$p_i^k = \frac{f\left(\theta_i^k\right)}{\sum_{i=1}^{n} f(\theta_i)} \quad (11)$$

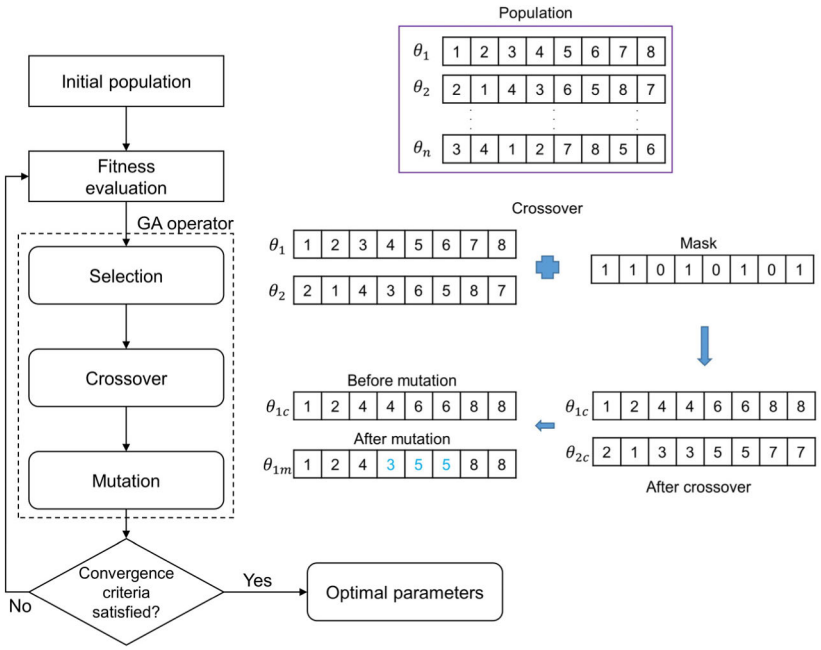

FIGURE 3. The genetic algorithm flowchart.

The next population is reproduced and formed through sample substitution, and each element corresponds to a reproducing probability p_i^k.

ii. *Crossover:* A new element called a child is generated and assigned to the population when genetic exchange among parents reaches a crossover point. A child is formulated as follows:

$$C = p_2^k + R_{cr}\left(p_1^k - p_2^k\right) \tag{12}$$

where R_{cr} denotes the ratio indicating the distance between child C and the better parent (p_1^k or p_2^k) and p_1^k is the parent with the best fitness score. The combination of elements in the next generation has a significant impact on GA performance, which is denoted by F_{cr}. This scheme is repeated for all elements of the population.

iii. *Mutation:* Some of the newly formed element genes can be mutated with low arbitrary probability P_m. This operator maintains the diversity in the population to avoid premature convergence and thereby increases the likelihood of generating improved elements.

Then, the GA sets the step to $k = k + 1$ and returns to the fitness evaluation stage. When the convergence conditions are met, the GA optimization process stops, and $\theta^* \equiv \arg max_{\theta_i^k} f\left(\theta_i^k\right)$ is produced as the optimum.

E. DIFFERENTIAL EVOLUTION ALGORITHM (DE)

First proposed by Storn and Price [53], the differential evolution (DE) algorithm is a population-based evolutionary scheme for optimizing fitness functions that are determined in continuous space. In this method, the generation of the population, the establishment of subsequent generations and fitness function evaluation are similar to the corresponding processes in the GA; however, crossover and mutation are implemented in different ways. Compared those in the GA, all the elements in the DE algorithm evolve. The evolved elements are directly exchanged among generations if the objective function scores are improved. Notably, global solutions can be produced by DE [43]. Fig. 4 shows the flowchart of the DE algorithm. The overview of this algorithm [54] is as follows.

1) INITIALIZATION OPERATOR

At step $k = 0$, the initial value of the j-th variable for the i-th element can be generated using the following equation:

$$\chi_{i,0}^j = \chi_{min}^j + rand\,(0,1) \cdot \left(\chi_{max}^j - \chi_{min}^j\right) \tag{13}$$

where $rand\,(0, 1)$ denotes a random number generator that generates values in the range of (0-1) from a uniform distribution and χ_{max}^j and χ_{min}^j are the lower and upper bound vectors of the variables. Then, a new mutant vector $\boldsymbol{V}_{i,g} = \left[v_{i,g}^j\right]\Big|_{i=1,j=1}^{p,n}$ is created for $\boldsymbol{x}_{i,g}$ by using the mutation operator, where p, n, and g are the number of elements, the number of optimization variables, and the generation index, respectively.

2) MUTATION OPERATOR

The new mutant vector is generated as:

$$\boldsymbol{V}_{i,g} = \boldsymbol{x}_{r1,g} + F \cdot \left(\boldsymbol{x}_{r2,g} - \boldsymbol{x}_{r3,g}\right) \tag{14}$$

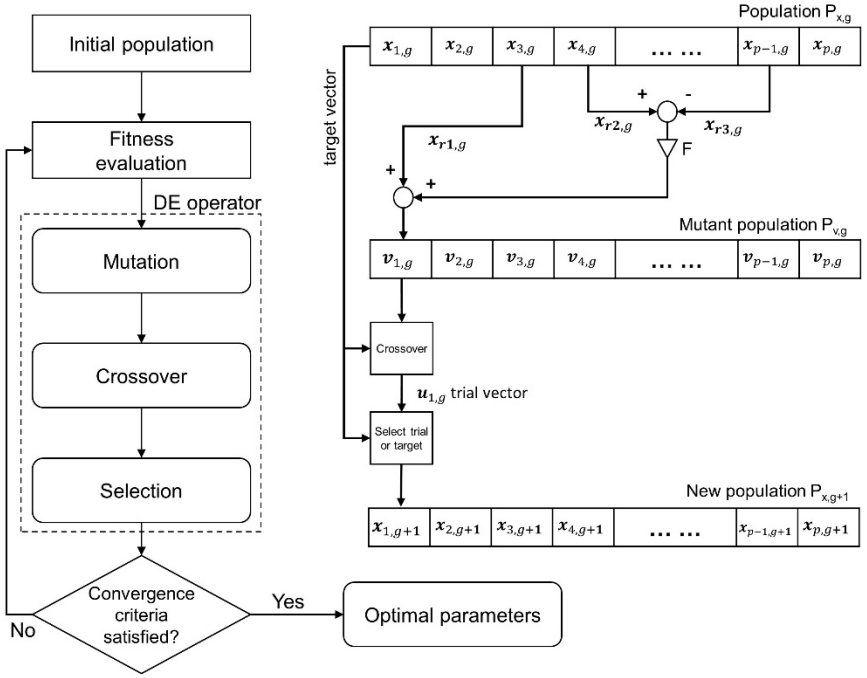

FIGURE 4. The differential evolution algorithm flowchart.

where $F \in (0, 1)$ denotes the scale factor, which controls population evolution; $r1$, $r2$, and $r3$ denote the randomly selected vector indexes.

3) CROSSOVER OPERATOR

A new element vector is generated based on combining the vectors $\boldsymbol{x}_{i,g}$ and $\boldsymbol{V}_{i,g}$.

$$\begin{aligned}\boldsymbol{u}_{i,g} &= \left[u_{i,g}^{j}\right]\Big|_{i=1,j=1}^{p,n} \\ &= \begin{cases} v_{i,g}^{j} & if\ \left(rand_j\,(0,1) \le C_r\ or\ j = j_{rand}\right) \\ \chi_{i,g}^{j} & otherwise \end{cases}\end{aligned} \tag{15}$$

where C_r denotes the crossover probability; j_{rand} is a randomly selected index; and the constraint $j = j_{rand}$ is used to guarantee that $\boldsymbol{u}_{i,g}$ includes at least one variable from $\boldsymbol{V}_{i,g}$.

4) SELECTION OPERATOR

This operator works to determine which candidate element is chosen for the next generation by comparing the trial vector $\boldsymbol{u}_{i,g}$ to the target vector $\boldsymbol{x}_{i,g}$.

$$\boldsymbol{x}_{i,g+1} = \begin{cases} \boldsymbol{u}_{i,g} & if\ f(\boldsymbol{u}_{i,g}) \le f(\boldsymbol{x}_{i,g}) \\ \boldsymbol{x}_{i,g} & otherwise \end{cases} \tag{16}$$

where f denotes the estimated objective function score. After the new population is generated, the operator processes are repeated until the optimal convergence or termination condition is met.

F. PERFORMANCE MEASURES

In this study, the performance of the models is evaluated based on various indicators, including the root mean square error (RMSE), correlation coefficient (CC), Nash-Sutcliffe efficiency (NSE), time lag (TL) between the predicted and observed occurrence of peak water levels, bias, mean absolute error (MAE), and mean absolute percentage error (MAPE). The indicators are formulated as follows.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(WL_i^{pre} - WL_i^{obs}\right)^2} \tag{17}$$

$$CC = \frac{\sum_{i=1}^{n}\left(WL_i^{pre} - WL_{mean}^{pre}\right)\left(WL_i^{obs} - WL_{mean}^{obs}\right)}{\sqrt{\sum_{i=1}^{n}\left(WL_i^{pre} - WL_{mean}^{pre}\right)^2\sum_{i=1}^{n}\left(WL_i^{obs} - WL_{mean}^{obs}\right)^2}} \tag{18}$$

$$NSE = 1 - \frac{\sum_{i=1}^{n}\left(WL_i^{obs} - WL_i^{pre}\right)^2}{\sum_{i=1}^{n}\left(WL_i^{obs} - WL_{mean}^{obs}\right)^2} \tag{19}$$

$$TL = T_{pre} - T_{obs} \tag{20}$$

$$BIAS = \frac{\sum_{i=1}^{n}\left(WL_i^{obs} - WL_i^{pre}\right)}{\sum_{i=1}^{n}\left(WL_i^{obs}\right)} \tag{21}$$

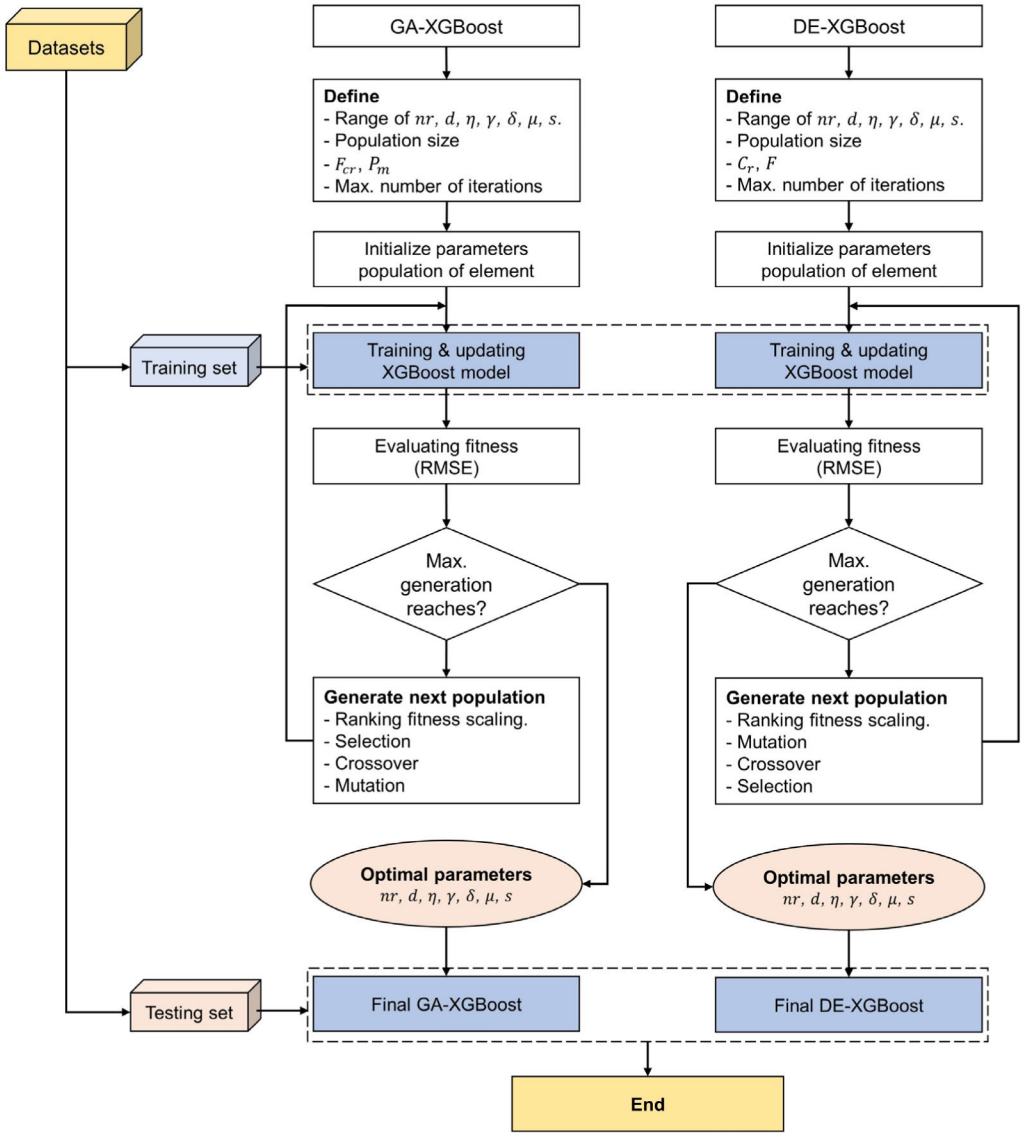

FIGURE 5. The GA-XGBoost and DE-XGBoost hybrid system.

$$MAE = \frac{\sum_{i=1}^{n} \left| WL_i^{obs} - WL_i^{pre} \right|}{n} \tag{22}$$

$$MAPE = \frac{\sum_{i=1}^{n} \left| \frac{WL_i^{obs} - WL_{mean}^{pre}}{WL_i^{obs}} \right|}{n} \times 100 \tag{23}$$

where WL_i^{pre} and WL_i^{obs} are the predicted and observed water levels at time step i, respectively; WL_{mean}^{obs} and WL_{mean}^{pre} are the average values of the observed and predicted water levels, respectively; and WL_{max}^{pre} and WL_{max}^{obs} are the predicted and observed peak water levels, respectively.

IV. MODEL IMPLEMENTATION

In this work, hybrid GA-XGBoost and DE-XGBoost models were developed based on R packages, including *xgboost* [30], [55] and *GA* [34], [56], to predict the multistep-ahead water level at Jungrang-gyo station in the Jungrang basin. In the developed system, the seven hyperparameters of the XGBoost model, including the maximum number of iterations (nr), maximum tree depth (d), learning rate (η), minimum loss reduction (γ), subsample ratio for columns (δ), minimum sum of instance weights (μ), and subsample ratio for training instances (s), were used to search the optimized sets by using the GA and DE algorithms. The meaning and range of the hyperparameters can be found in the above-mentioned articles. In this study, the values of nr, d, η, γ, δ, μ, and s were set in the ranges of [50-800], [3-10], [0.005-0.3], [0-3], [0.3-0.8], [0-10], and [0.4-1], respectively. The two developed hybrid systems in which RMSE was used as an objective function of the optimization process are shown in Fig. 5. RMSE is widely used in hydrology, forecasting, and regression analysis to determine how concentrated the predicted and observed data are around the

line of best fit. This indicator can be sensitive to the peak values of water levels that need to be detected by the models.

For comparison, CART and RF models were developed based on R packages, including *rpart* [57] and *randomForest* [48], [58]. For the CART model, the complexity parameter was tuned in the range of (0, 1) with the grid search method. For the RF model, the number of trees (*ntree*) and the randomly split samples of variables (*mtry*) have the main influence on the modeling results. As suggested in the package documentation and literature [27], [59], the *ntree* parameter for an RF should be sufficiently large. The parameters *mtry* and *ntree* were tuned in the ranges of [1-15] and [1000-2000], respectively, with the grid search method.

The k-fold cross-validation method was applied to mitigate the overfitting problem in the tree-based models. All the models (CART, RF, GA-XGBoost, and DE-XGBoost) used 5-fold cross-validation, which has been widely used in previous studies [23]; this approach is shown in Fig. 6(a). Each hybrid model is described in detail in the following paragraphs.

For the GA-XGBoost model, the population size (P_s) and the maximum number of generations (G_{max}) are the most sensitive parameters in the GA and were determined based on a parametric test. The crossover probability F_{cr} between pairs of chromosomes should be a large value (70%-80%). The mutation probability P_m for a parent chromosome should be a small value because the probability of mutation occurring is small. Therefore, in this study, F_{cr} and P_m in the GA were selected to be 70% and 3%, respectively, and were set at the initial step. In this study, the uniform crossover method was applied; this approach utilizes a mixing rate to copy a gene from each parent to create child genes (Fig. 3). To set suitable values of G_{max} and P_s, nine GA-XGBoost models that used the 5-fold cross-validation technique were developed based on a set of P_s values, including 20, 50, 100, 150, 200, 250, 300, 400, and 500. Fig. 6(b) presents the GA-XGBoost performance and indicates that the RMSE values were stable after 700 generations and that the best P_s was 200 based on the lowest RMSE. Similar to those in the GA-XGBoost model, the mutation probability (F) and crossover probability (C_r) for the DE operator were chosen as 3% and 70%, respectively. As illustrated in Fig. 6(c), the RMSE values of DE-XGBoost were unchanged after 320 generations, and the optimal population size was 150. For the stopping criteria in these models, the maximum number of iterations (*maxiter*) was selected as 1000, and the consecutive number of iterations without an improvement in the fitness score (*run*) was chosen as 100. Notably, the abovementioned optimization process was for 1-hour-ahead water level prediction. For the remaining lead times, the parameters of the optimization models were set to the same values as those in the 1-hour case.

V. RESULTS AND DISCUSSION

This section presents the prediction performance results for the hybrid models (GA-XGBoost and DE-XGBoost) and other tree-based models (RF and CART) for six test events. The detailed results and discussion are as follows.

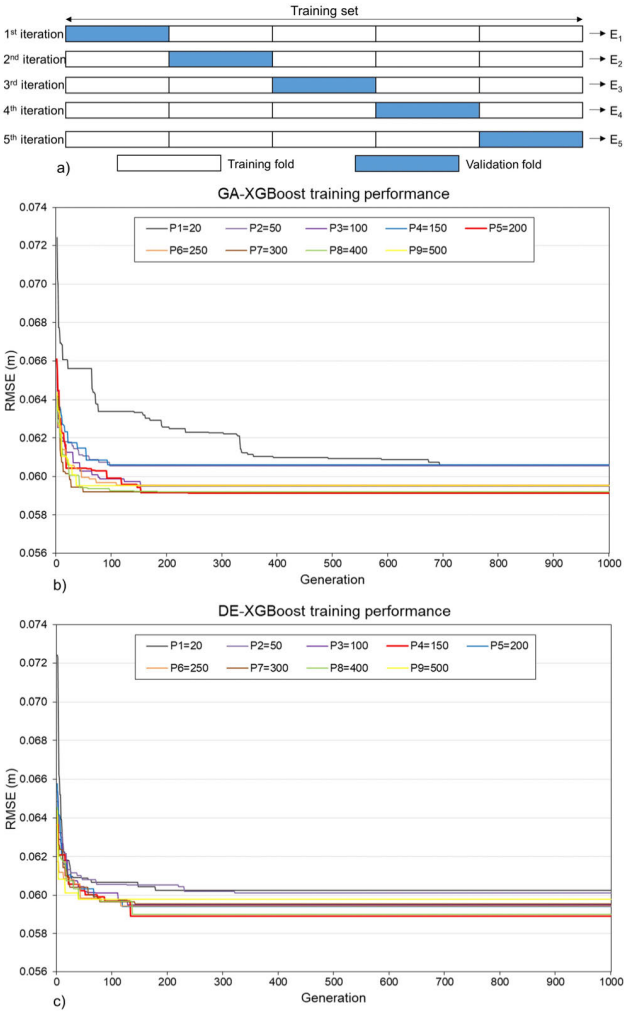

FIGURE 6. Construction of GA-XGBoost and DE-XGBoost hybrid models: a) k-fold cross-validation approach, b) GA-XGBoost performance, and c) DE-XGBoost performance.

Table 3 summarizes a performance comparison of the four tree-based models—GA-XGBoost, DE-XGBoost, RF, and CART—regarding the indicators RMSE, CC, NSE, MA, and MAPE for one- to six-step-ahead predictions of water level. In addition to comparing the abovementioned indicators for the different models, the complexity of the models was evaluated. The Akaike information criterion (AIC) was used to examine model complexity [60]. The AIC formula is as follows:

$$AIC = n \times log\left(\sqrt{RMSE}\right) + 2k \tag{24}$$

where n denotes the sample number and k denotes the number of leaf nodes corresponding to rules based on certain predictors.

Table 3 indicates that the accuracy of the prediction models generally decreased as the number of time steps increased. For the 1-hour-ahead forecast, all the tree-based models

TABLE 3. Performance of different tree-based models in one- to six-step-ahead water level prediction for test events.

Time steps	Indicators	GA-XGBoost	DE-XGBoost	RF	CART
1	RMSE	0.0637	0.0662	0.0733	0.0788
	CC	0.9873	0.9867	0.9831	0.9806
	NSE	0.9743	0.9723	0.9661	0.9607
	MAE	0.0232	0.0283	0.0335	0.0252
	MAPE	2.18	2.27	3.76	2.99
	AIC	9539.9	6769.0	59257.6	1600.4
2	RMSE	0.1117	0.1091	0.1164	0.1298
	CC	0.9607	0.9625	0.9565	0.9477
	NSE	0.9210	0.9248	0.9144	0.8934
	MAE	0.0440	0.0438	0.0519	0.0476
	MAPE	4.11	4.15	5.56	4.27
	AIC	21263.6	13626.7	59389.0	2770.1
3	RMSE	0.1445	0.1446	0.1490	0.1624
	CC	0.9325	0.9322	0.9273	0.9165
	NSE	0.8678	0.8677	0.8595	0.8330
	MAE	0.0579	0.0575	0.0652	0.0643
	MAPE	5.35	5.27	6.68	6.00
	AIC	14138.6	8250.7	59459.3	-242.2
4	RMSE	0.1778	0.1794	0.1815	0.1908
	CC	0.8948	0.8930	0.8897	0.8792
	NSE	0.7998	0.7961	0.7914	0.7694
	MAE	0.0713	0.0751	0.0790	0.0842
	MAPE	6.49	6.96	7.81	8.33
	AIC	26239.4	40762.0	59515.3	-332.4
5	RMSE	0.2078	0.2077	0.2114	0.2258
	CC	0.8526	0.8529	0.8468	0.8253
	NSE	0.7264	0.7266	0.7169	0.6769
	MAE	0.0862	0.0880	0.0930	0.1065
	MAPE	7.85	8.14	9.11	10.73
	AIC	21005.7	12273.6	59558.6	-360.6
6	RMSE	0.2330	0.2351	0.2362	0.2474
	CC	0.8106	0.8065	0.8041	0.7881
	NSE	0.6559	0.6497	0.6463	0.6119
	MAE	0.0990	0.1011	0.1059	0.1190
	MAPE	9.21	9.46	10.41	11.88
	AIC	9762.2	16764.7	59590.1	-326.7

produced good results, with high CC and NSE values and low RMSE and MAE values. For the 2-hour-ahead prediction, the performance of the CART declined substantially compared to that in the 1-hour-ahead prediction. In detail, the NSE and CC values of the CART approach decreased from 0.9607 and 0.9806 to 0.8934 and 0.9477, respectively. Additionally, the RMSE and MAE values of the CART increased from 0.0788 and 0.0252 to 0.1298 and 0.0476, respectively. As the time step length increased, the CART accuracy continually decreased. The RF performed better than the CART. For the six-hour ahead prediction, the RF displayed good performance, with slightly lower NSE and CC values and higher RMSE and MAE values compared to those of DE-XGBoost. GA- and DE-XGBoost outperformed the RF and CART models in terms of all the performance indicators. For the 1-, 4-, and 6-hour-ahead predictions, GA-XGBoost exhibited slightly better performance than DE-XGBoost; however, the opposite result was observed the 2-hour-ahead forecast. For the 3- and 5-hour-ahead predictions, both models displayed similar performance. The relative error of the hybrid GA-XGBoost model (MAPE = 2.18%) was approximately 1.58% lower than that of the RF and 0.81% lower than that of the CART for the 1-hour-ahead prediction. For the rest of the time steps, both hybrid models exhibited lower MAPE values than the RF and CART models. The MAPE values of GA-XGBoost were slightly lower than those of DE-XGBoost for five of the six time steps. For the AIC indicator, which considers both model complexity and accuracy, the AIC value for the GA-XGBoost model (9539.9) was approximately one-sixth the AIC value of the RF in the 1-hour ahead-prediction. The AIC values of the RF for other time steps were the highest among those of all models. RF performance in terms of RMSE, CC, NSE, MAE, and MAPE was less reliable than that of the XGBoost models, even though it had a more complex model structure. The AIC values of DE-XGBoost were slightly lower than those of GA-XGBoost for four of the six time steps. The CART exhibited the smallest AIC values for all time steps due to the simple structure (a single tree) of the algorithm. However, the performance of the CART was not reliable, as shown in Table 3 and the following analyses.

To further investigate the performance of the GA-XGBoost, DE-XGBoost, RF, and CART models, continuous 6-hour lead time predictions at each time step t were obtained. Notably, the results of multistep-ahead models were combined into a 6-hour forecasting series at each time step t. Fig. 7 displays the performance of the models for continuous 6-hour lead time predictions for the six test events. The performance criteria were based on the mean 6-hour predictions generated at each time t for the test events. As shown in Fig. 7, the two hybrid models outperformed the other models, with lower RMSE and MAE values and higher CC and NSE values in 6-hour lead time predictions. Notably, DE-XGBoost exhibited better performance than GA-XGBoost in terms of the RMSE and MAE, and both models yielded similar CC and NSE results. As in the multistep-ahead predictions, the RF performed better than the CART in terms of RMSE, CC and MAE. For NSE results, the CART model displayed slightly better accuracy than the RF model for 1-hour to 4-hour lead time and showed approximately the same as both hybrid models for the 1-hour and 4-hour lead time. However, this single phenomenon does not suggest that the performance of the CART is robust when comprehensively considering all the performance features.

Fig. 8 provides an intuitive way to evaluate the performance of the hybrid tree-based, RF, and CART models in terms of hydrographs of the observed versus one-, three-, and five-step-ahead predicted water levels for four events: 37, 40, 41, and 42. For the one-hour-ahead forecast, all the models captured the behavior of the time-series water level well. Nevertheless, CART displayed instability in predicting water level variations (Fig. 8(d), (g), and (j)). The RF exhibited generally underestimated forecasts when the water level decreased (Fig. 8(g)) and a larger time lag (Fig. 8(j)) than the two hybrid models. In addition, the RF produced a notable underestimation of the second peak water level compared to the two XGBoost-based models for event 37 (Fig. 8(a)). In long step-ahead predictions, the CART displayed poor

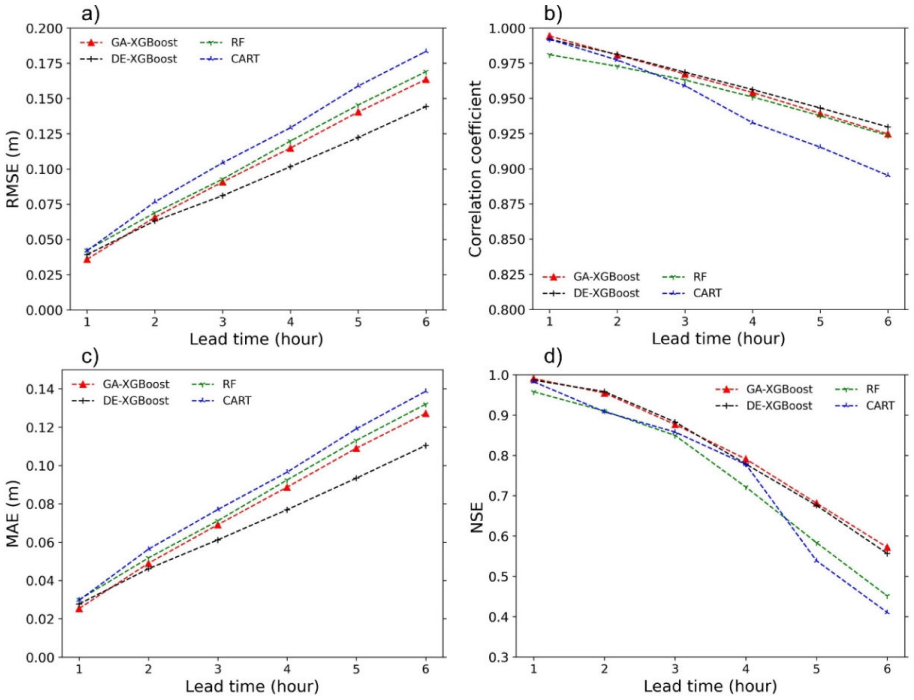

FIGURE 7. The performance of the models in terms of the RMSE, correlation coefficient, NSE, and MAE for continuous 6-hour lead time predictions based on the test event.

performance in capturing the behavior of water levels during the events (Fig. 8(c), (e-f), (h-i), and (k-l)). Compared to the GA- and DE-XGBoost models, the RF exhibited more underestimated predictions of the peak water level for events 37, 40, and 42 (Fig. 8(b), (e-f), and (k-l)). The GA- and DE-XGBoost predictions illustrated very similar behavior for the various prediction steps, and there was only small difference in water level variations in both models (Fig. 8(g-i) and (k-l)).

TABLE 4. Bias performance of the models for the test events.

Time steps	Bias			
	GA-XGBoost	DE-XGBoost	RF	CART
1	-0.00074	-0.00405	0.00562	-0.00308
2	0.00383	0.00588	0.02663	-0.00305
3	0.01850	0.02083	0.04251	0.00990
4	0.04071	0.04074	0.05859	0.04348
5	0.06841	0.06671	0.07703	0.06545
6	0.08190	0.09473	0.09618	0.06885

Table 4 shows the bias performance of the models for the test events. The CART displayed good bias results for the prediction steps; however, this result did not indicate overall good performance. The two hybrid models outperformed the RF for all multistep-ahead predictions. GA-XGBoost exhibited slightly better performance than DE-XGBoost for four of the six step-ahead predictions.

Table 5 presents the time lag between the predicted and observed occurrences of the water level peaks for the test events. The hybrid models GA- and DE-XGBoost outperformed the RF and CART, with lower time lags for the water level peak in the multistep-ahead predictions for events 39, 40, and 42. In addition, both hybrid models yielded lower time lags than that of the RF for the 4-hour- and 6-hour-ahead predictions for events 37 and 41, respectively. The RF, GA-XGBoost and DE-XGBoost models displayed similar time larges for events 37, 38, and 41. Generally, the two hybrid models can not only produce more accurate predictions for single-peak events but also yield better performance for multipeak events compared to the RF and CART.

Fig. 9 shows the scatter plots of the GA-XGBoost, DE-XGBoost, RF and CART models for the $t + 3$ and $t + 5$ predictions. As shown in Fig. 9, the two hybrid models yielded better performance (with higher CC values and narrowly dispersed points for $t + 3$ predictions (Fig. 9(a)) and a more reasonable distribution of points for $t + 5$ predictions (Fig. 9(b)) than the CART. The RF outperformed the CART; however, compared to the hybrid-XGBoost models, more underestimated points and a lower CC value were produced. The points of the CART are generally scattered in layers. For longer step-ahead predictions (e.g., $t + 5$), the differentiation among point layers is clearer than that for $t + 3$ predictions. This finding suggests that as a simple tree-based model, the CART is not suitable for multistep-ahead forecasting.

To further compare these four tree-based models in an intuitive way, a Taylor diagram was created. A Taylor diagram is a mathematically based graphical diagram that provides a visualized representation of how closely predictions and observations are based on the correlation coefficient, RMSE,

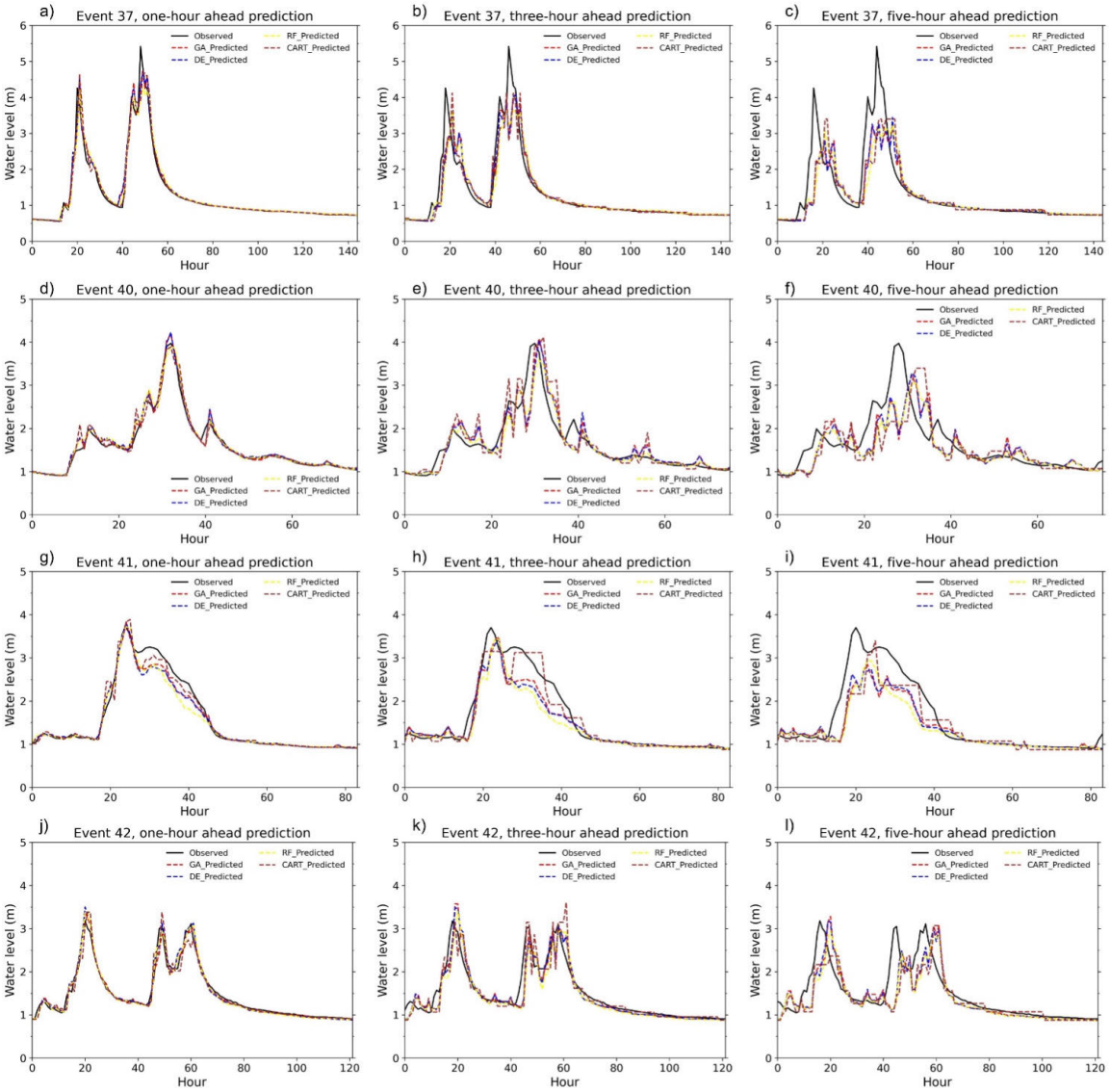

FIGURE 8. Time series of model-predicted water levels for t + 1, t + 3, and t + 5 step-ahead cases.

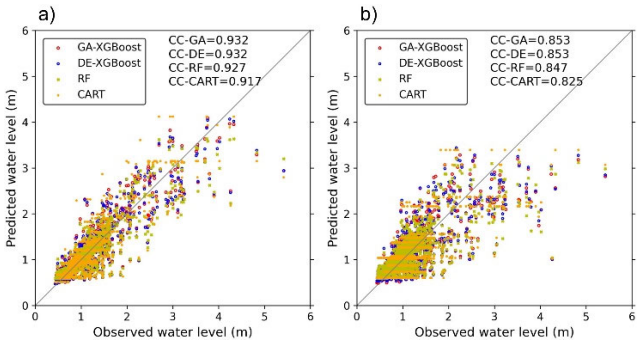

FIGURE 9. Scatter plots of the GA-XGBoost, DE-XGBoost, RF, and CART models for a) t + 3 and b) t + 5 predictions.

and standard deviation. Fig. 10 shows the Taylor diagram of the multistep-ahead predictions ($t+1$, $t+2$, $t+3$, and $t+5$) for the test events. As shown, GA-XGBoost and DE-XGBoost produce the results close to the observations, with higher CC and lower RMSE values than those of the other models for all time steps.

Although the hybrid XGBoost models outperformed the RF and CART models, the GA- and DE-XGBoost models display some limitations that need to be discussed. The two hybrid models produced underestimations of water level predictions (Fig. 8(c), (f), and (i) and Fig. 9 (b)) and large time lags for peak water levels (Table 5) for 5-hour- and 6-hour-ahead predictions. The models displayed relatively poor performance in terms of NSE for long prediction steps. Specifically, the NSE values were approximately 0.65 for the 6-hour-ahead predictions (Table 3) and were lower than 0.6 for continuous 6-hour lead time predictions (Fig. 7(d)). This decrease is performance might be related to the characteristics of the basin. The examined basin is relatively small (299.87 km^2), and the water level and streamflow are very sensitive to hydrological variables, especially rainfall variability. Obtaining reliable rainfall predictions for long

TABLE 5. Performance of the models based on the time lag of the peak water levels for the test events.

Models	Events	Time lag of the peak water level (hour)					
		t+1	t+2	t+3	t+4	t+5	t+6
GA-XGBoost	37	1; 1	2; 1	3; 3	4; 3	5, 8	6; 8
	38	-1	0	1	2	3	4
	39	-1	1	0	1	2	3
	40	0	0	1	2	3	5
	41	0	1	2	1	4	3
	42	0; 0; 1	1; 1; -1	1; 0; 0	2; 1; 4	4; 2; 5	5; 3; 6
DE-XGBoost	37	1; 1	2; 1	3; 3	4; 3	5; 8	4; 8
	38	-1	0	1	2	3	4
	39	1	2	0	1	3	3
	40	0	0	1	2	3	4
	41	0	1	2	1	4	3
	42	0; 0; 1	1; 1; -1	1; 0; 0	3; 1; 4	3; 2; 3	5; 3; 6
RF	37	1; 1	2; 1	3, 3	4; 6	5; 7	6; 8
	38	-1	0	1	2	3	4
	39	1	2	3	4	5	6
	40	0	1	1	2	4	5
	41	0	1	2	2	3	5
	42	1; 0; 1	1; 1; 1	2; 0; 1	3; 1; 3	4; 2; 3	5; 3; 4
CART	37	1; 1	2; 4	3; 2	3; 3	5; 6	6; 7
	38	-1	0	1	2	3	4
	39	-1	2	3	4	3	4
	40	0	-1	2	2	4	5
	41	1	1	-2	-1	5	7
	42	1; 0; 1	1; -2; 1	1; -1; 3	3; 1; 4	4; 5; 5	5; 3; 4

Note: Multiple peak water levels for an event are separated by semicolons. For example, event 42 had three peaks, and the time lags obtained with the GA-XGBoost model were "0; 0; 1".

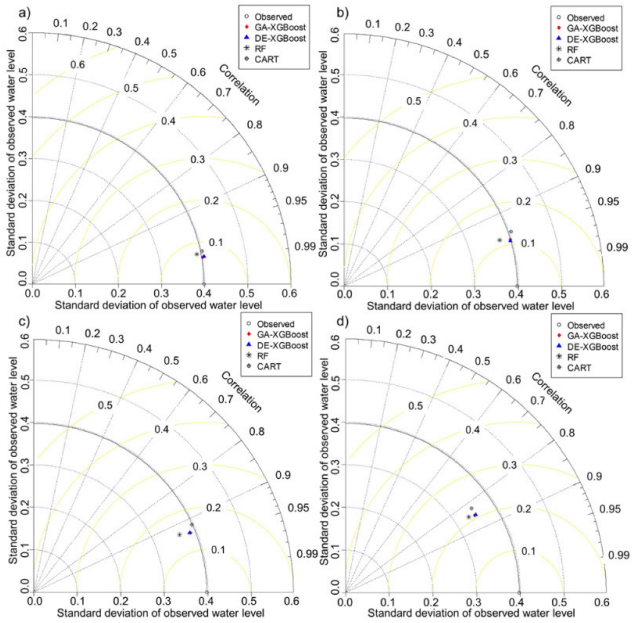

FIGURE 10. Taylor diagrams of the multistep-ahead models: (a) t + 1, (b) t + 2, (c) t + 3, and (d) t + 5.

steps (5 and 6 hours ahead) remains a challenge for forecasting systems. The effectiveness of rainfall forecasting often deteriorates as the prediction step increases. In addition, using past data without considering the variations in other hydrological features and other correlated parameters, such as basin cover changes over multiple decades, radiation, and wind speed, might be a source of error. Therefore, to improve the accuracy of the models, it might be necessary to add some related basin features and other correlated predictors to capture the changes in the basin over long periods. When considering model complexity and accuracy together, the hybrid XGBoost models yield high AIC values. This result reflects the trade-off of using models with complex tree structures to obtain high prediction accuracy.

VI. CONCLUSION

In the context of effective urban flood control, it is necessary to predict reliable water levels during heavy rainfall events. Accurate water level prediction remains a scientific challenge and has received considerable attention due to the nonlinear and nonstationary nature of water levels. In this study, we developed and examined four tree-based models: GA-XGBoost, DE-XGBoost, RF, and CART models. Through a case study in the Jungrang urban basin, South Korea, the performance of the four tree-based models was compared based on multistep-ahead predictions of water levels. The data were collected from rain gauges and water level stations from 2003 to 2020, with 42 heavy rain events during that span.

The obtained results showed that (1) the two hybrid models, GA-XGBoost and DE-XGBoost, outperformed the RF and CART models in multistep-ahead water level prediction based on the considered performance measures, including lower RMSE and MAE values and higher CC and NSE values. In addition, compared to the benchmark models, the hybrid XGBoost models provided better time lag results for water level peaks and superior bias. (2) Generally, the performance of the GA-XGBoost and DE-XGBoost models was similar; however, there were small differences between them. DE-XGBoost provided better accuracy than GA-XGBoost for continuous 6-hour lead time prediction based on values generated at every time step. GA-XGBoost performed slightly better than DE-XGBoost for series of separate time step predictions. (3) The CART, as a simple tree-based model, was not suitable for multistep-ahead water level predictions.

Although there are some limitations when considering long step-ahead predictions (i.e., the underestimation of water levels, large time lags for water level peaks, and poor NSE measures), the hybrid XGBoost model might be a superior option to the existing models used for hourly water level prediction. In future work, tree-based models should be explored, and their performance should be assessed at different basin scales and temporal scales. To improve hourly multistep predictions, an approach that combines monthly forecasting, periodic patterns and daily and hourly forecasting should be investigated. In addition, the comparison of the tree-based models with other machine learning methods and the optimization of input predictors should be examined in future research.

REFERENCES

[1] D. Hine and J. W. Hall, "Information gap analysis of flood model uncertainties and regional frequency analysis," *Water Resour. Res.*, vol. 46, no. 1, pp. 1–18, Jan. 2010.

[2] V. Aich, S. Liersch, T. Vetter, S. Fournet, J. C. M. Andersson, S. Calmanti, F. H. A. van Weert, F. F. Hattermann, and E. N. Paton, "Flood projections within the Niger River Basin under future land use and climate change," *Sci. Total Environ.*, vol. 562, pp. 666–677, Aug. 2016.

[3] X. Wang, G. Kinsland, D. Poudel, and A. Fenech, "Urban flood prediction under heavy precipitation," *J. Hydrol.*, vol. 577, Oct. 2019, Art. no. 123984.

[4] S. Sharma, R. Siddique, S. Reed, P. Ahnert, and A. Mejia, "Hydrological model diversity enhances streamflow forecast skill at short-to medium-range timescales," *Water Resour. Res.*, vol. 55, no. 2, pp. 1510–1530, Feb. 2019.

[5] D. B. Wijayarathne and P. Coulibaly, "Identification of hydrological models for operational flood forecasting in St. John's, Newfoundland, Canada," *J. Hydrol., Regional Stud.*, vol. 27, Feb. 2020, Art. no. 100646.

[6] D. H. Nguyen and D.-H. Bae, "Correcting mean areal precipitation forecasts to improve urban flooding predictions by using long short-term memory network," *J. Hydrol.*, vol. 584, May 2020, Art. no. 124710.

[7] Y. Zhou, S. Guo, and F.-J. Chang, "Explore an evolutionary recurrent ANFIS for modelling multi-step-ahead flood forecasts," *J. Hydrol.*, vol. 570, pp. 343–355, Mar. 2019.

[8] R. M. Adnan, Z. Liang, S. Trajkovic, M. Zounemat-Kermani, B. Li, and O. Kisi, "Daily streamflow prediction using optimally pruned extreme learning machine," *J. Hydrol.*, vol. 577, Oct. 2019, Art. no. 123981.

[9] E. Meng, S. Huang, Q. Huang, W. Fang, L. Wu, and L. Wang, "A robust method for non-stationary streamflow prediction based on improved EMD-SVM model," *J. Hydrol.*, vol. 568, pp. 462–478, Jan. 2019.

[10] R. Noori, A. R. Karbassi, A. Moghaddamnia, D. Han, M. H. Zokaei-Ashtiani, A. Farokhnia, and M. G. Gousheh, "Assessment of input variables determination on the SVM model performance using PCA, Gamma test, and forward selection techniques for monthly stream flow prediction," *J. Hydrol.*, vol. 401, nos. 3–4, pp. 177–189, 2011.

[11] F.-J. Chang, P.-A. Chen, Y.-R. Lu, E. Huang, and K.-Y. Chang, "Real-time multi-step-ahead water level forecasting by recurrent neural networks for urban flood control," *J. Hydrol.*, vol. 517, pp. 836–846, Dec. 2014.

[12] T. Zhang and Z. W. Geem, "Review of harmony search with respect to algorithm structure," *Swarm Evol. Comput.*, vol. 48, pp. 31–43, Aug. 2019.

[13] D. Liu, W. Jiang, L. Mu, and S. Wang, "Streamflow prediction using deep learning neural network: Case study of Yangtze River," *IEEE Access*, vol. 8, pp. 90069–90086, 2020.

[14] D. H. Nguyen, J.-B. Kim, and D.-H. Bae, "Improving radar-based rainfall forecasts by long short-term memory network in urban basins," *Water*, vol. 13, no. 6, p. 776, Mar. 2021.

[15] X.-H. Le, D.-H. Nguyen, S. Jung, M. Yeon, and G. Lee, "Comparison of deep learning techniques for river streamflow forecasting," *IEEE Access*, vol. 9, pp. 71805–71820, 2021.

[16] M. Cheng, F. Fang, T. Kinouchi, I. M. Navon, and C. C. Pain, "Long lead-time daily and monthly streamflow forecasting using machine learning methods," *J. Hydrol.*, vol. 590, Nov. 2020, Art. no. 125376.

[17] A. Osman, H. A. Afan, M. F. Allawi, O. Jaafar, A. Noureldin, F. M. Hamzah, A. N. Ahmed, and A. El-shafie, "Adaptive fast orthogonal search (FOS) algorithm for forecasting streamflow," *J. Hydrol.*, vol. 586, Jul. 2020, Art. no. 124896.

[18] M. B. Wagena, D. Goering, A. S. Collick, E. Bock, D. R. Fuka, A. Buda, and Z. M. Easton, "Comparison of short-term streamflow forecasting using stochastic time series, neural networks, process-based, and Bayesian models," *Environ. Model. Softw.*, vol. 126, Apr. 2020, Art. no. 104669.

[19] A. D. Mehr, E. Kahya, and E. Olyaie, "Streamflow prediction using linear genetic programming in comparison with a neuro-wavelet technique," *J. Hydrol.*, vol. 505, pp. 240–249, Nov. 2013.

[20] A. D. Mehr, "An improved gene expression programming model for streamflow forecasting in intermittent streams," *J. Hydrol.*, vol. 563, pp. 669–678, Aug. 2018.

[21] S. J. Hadi and M. Tombul, "Monthly streamflow forecasting using continuous wavelet and multi-gene genetic programming combination," *J. Hydrol.*, vol. 561, pp. 674–687, Jun. 2018.

[22] Z. M. Yaseen, I. Ebtehaj, H. Bonakdari, R. C. Deo, A. D. Mehr, W. H. M. W. Mohtar, L. Diop, A. El-Shafie, and V. P. Singh, "Novel approach for streamflow forecasting using a hybrid ANFIS-FFA model," *J. Hydrol.*, vol. 554, pp. 263–276, Nov. 2017.

[23] L. Ni, D. Wang, J. Wu, Y. Wang, Y. Tao, J. Zhang, and J. Liu, "Streamflow forecasting using extreme gradient boosting model coupled with Gaussian mixture model," *J. Hydrol.*, vol. 586, Jul. 2020, Art. no. 124901.

[24] S. Gharaei-Manesh, A. Fathzadeh, and R. Taghizadeh-Mehrjardi, "Comparison of artificial neural network and decision tree models in estimating spatial distribution of snow depth in a semi-arid region of Iran," *Cold Regions Sci. Technol.*, vol. 122, pp. 26–35, Feb. 2016.

[25] E. Lee and S. Kim, "Characterization of runoff generation in a mountainous hillslope according to multiple threshold behavior and hysteretic loop features," *J. Hydrol.*, vol. 590, Nov. 2020, Art. no. 125534.

[26] H. I. Erdal and O. Karakurt, "Advancing monthly streamflow prediction accuracy of CART models using ensemble learning paradigms," *J. Hydrol.*, vol. 477, pp. 119–128, Jan. 2013.

[27] L. Schoppa, M. Disse, and S. Bachmair, "Evaluating the performance of random forest for large-scale flood discharge simulation," *J. Hydrol.*, vol. 590, Nov. 2020, Art. no. 125531.

[28] M. Abbasi, A. Farokhnia, M. Bahreinimotlagh, and R. Roozbahani, "A hybrid of random forest and deep auto-encoder with support vector regression methods for accuracy improvement and uncertainty reduction of long-term streamflow prediction," *J. Hydrol.*, vol. 597, Jun. 2021, Art. no. 125717.

[29] M. Li, Y. Zhang, J. Wallace, and E. Campbell, "Estimating annual runoff in response to forest change: A statistical method based on random forest," *J. Hydrol.*, vol. 589, Oct. 2020, Art. no. 125168.

[30] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 785–794.

[31] S. J. Hadi, M. Tombul, S. Q. Salih, N. Al-Ansari, and Z. M. Yaseen, "The capacity of the hybridizing wavelet transformation approach with data-driven models for modeling monthly-scale streamflow," *IEEE Access*, vol. 8, pp. 101993–102006, 2020.

[32] H. R. Madvar, M. Dehghani, R. Memarzadeh, E. Salwana, A. Mosavi, and S. Shahab, "Derivation of optimized equations for estimation of dispersion coefficient in natural streams using hybridized ANN with PSO and CSO algorithms," *IEEE Access*, vol. 8, pp. 156582–156599, 2020.

[33] J. H. Holland, *Adaptation in Natural and Artificial Systems: An Introductory Analysis With Applications to Biology, Control, and Artificial Intelligence*. Cambridge, MA, USA: MIT Press, 1992.

[34] L. Scrucca, "GA: A package for genetic algorithms in R," *J. Statist. Softw.*, vol. 53, no. 4, pp. 1–37, 2013.

[35] N.-V. Luat, J. Shin, and K. Lee, "Hybrid BART-based models optimized by nature-inspired metaheuristics to predict ultimate axial capacity of CCFST columns," *Eng. Comput.*, pp. 1–30, Jul. 2020.

[36] E. Dodangeh, M. Panahi, F. Rezaie, S. Lee, D. Tien Bui, C.-W. Lee, and B. Pradhan, "Novel hybrid intelligence models for flood-susceptibility prediction: Meta optimization of the GMDH and SVR models with the genetic algorithm and harmony search," *J. Hydrol.*, vol. 590, Nov. 2020, Art. no. 125423.

[37] P. Horton, M. Jaboyedoff, and C. Obled, "Using genetic algorithms to optimize the analogue method for precipitation prediction in the Swiss Alps," *J. Hydrol.*, vol. 556, pp. 1220–1231, Jan. 2018.

[38] M. Chlumecký, J. Buchtele, and K. Richta, "Application of random number generators in genetic algorithms to improve rainfall-runoff modelling," *J. Hydrol.*, vol. 553, pp. 350–355, Oct. 2017.

[39] Z. A. Al-Sudani, S. Q. Salih, A. Sharafati, and Z. M. Yaseen, "Development of multivariate adaptive regression spline integrated with differential evolution model for streamflow simulation," *J. Hydrol.*, vol. 573, pp. 1–12, Jun. 2019.

[40] H. R. Pourghasemi, S. V. Razavi-Termeh, N. Kariminejad, H. Hong, and W. Chen, "An assessment of Metaheuristic approaches for flood assessment," *J. Hydrol.*, vol. 582, Mar. 2020, Art. no. 124536.

[41] A. Gholami, H. Bonakdari, I. Ebtehaj, B. Gharabaghi, S. R. Khodashenas, S. H. A. Talesh, and A. Jamali, "A methodological approach of predicting threshold channel bank profile by multi-objective evolutionary optimization of ANFIS," *Eng. Geol.*, vol. 239, pp. 298–309, May 2018.

[42] H. Azimi, H. Bonakdari, I. Ebtehaj, S. H. Ashraf Talesh, D. G. Michelson, and A. Jamali, "Evolutionary Pareto optimization of an ANFIS network for modeling scour at pile groups in clear water condition," *Fuzzy Sets Syst.*, vol. 319, pp. 50–69, Jul. 2017.

[43] A. Elçi and M. T. Ayvaz, "Differential-evolution algorithm based optimization for the site selection of groundwater production wells with the consideration of the vulnerability concept," *J. Hydrol.*, vol. 511, pp. 736–749, Apr. 2014.

[44] Q. Wang, R. Liu, C. Men, L. Guo, and Y. Miao, "Temporal-spatial analysis of water environmental capacity based on the couple of SWAT model and differential evolution algorithm," *J. Hydrol.*, vol. 569, pp. 155–166, Feb. 2019.
[45] D. Kurtzman, S. Navon, and E. Morin, "Improving interpolation of daily precipitation for hydrologic modelling: Spatial patterns of preferred interpolators," *Hydrol. Processes*, vol. 23, no. 23, pp. 3281–3291, Nov. 2009.
[46] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, *Classification and Regression Trees*. New York, NY, USA: Chapman & Hall, 1984.
[47] J. Han, K. Mao, T. Xu, J. Guo, Z. Zuo, and C. Gao, "A soil moisture estimation framework based on the CART algorithm and its application in China," *J. Hydrol.*, vol. 563, pp. 65–75, Aug. 2018.
[48] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[49] C. Carranza, C. Nolet, M. Pezij, and M. van der Ploeg, "Root zone soil moisture estimation with random forest," *J. Hydrol.*, vol. 593, Feb. 2021, Art. no. 125840.
[50] J. H. Friedman, "Stochastic gradient boosting," *Comput. Statist. Data Anal.*, vol. 38, no. 4, pp. 367–378, 2002.
[51] J. H. Holland, *Adaptation in Natural and Artifical Systems: An Introductory Analysis With Applications to Biology, Control, and Artificial Intelligence*. Cambridge, MA, USA: MIT Press, 1975.
[52] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*. Reading, MA, USA: Addison-Wesley, 1989.
[53] R. Storn and K. Price, "Differential evolution—A simple and efficient heuristic for global optimization over continuous spaces," *J. Global Optim.*, vol. 11, no. 4, pp. 341–359, 1997.
[54] M. Salomon, G.-R. Perrin, F. Heitz, and J.-P. Armspach, *Differential Evolution: A Practical Approach to Global Optimization*. Berlin, Germany: Springer-Verlag, 2006.
[55] T. Chen and T. He, "Xgboost: Extreme gradient boosting," *R Packag. Version 1.2.0.1*, pp. 1–4, 2020.
[56] L. Scrucca, "GA: Genetic algorithm," *R Packag. Version 3.2*, pp. 1–44, 2019.
[57] T. Therneau, B. Atkinson, and B. Ripley, "rpart: Recursive partitioning for classification, regression and survival trees," *R Package Version 4.1-15*, pp. 1–4, 2019. [Online]. Available: https://cran.r-project.org/package=rpart
[58] A. Liaw and M. Wiener, "randomForest," *R Package Version*, vol. 4, pp. 6–14, 2018.
[59] L. Gudmundsson and S. I. Seneviratne, "Observation-based gridded runoff estimates for Europe (E-RUN version 1.1)," *Earth Syst. Sci. Data*, vol. 8, no. 2, pp. 279–295, Jul. 2016.
[60] H. Bonakdari, H. Moeeni, I. Ebtehaj, M. Zeynoddin, A. Mahoammadian, and B. Gharabaghi, "New insights into soil temperature time series modeling: Linear or nonlinear?" *Theor. Appl. Climatol.*, vol. 135, nos. 3–4, pp. 1157–1177, 2019.

DUC HAI NGUYEN received the B.Sc. degree from Thuyloi University, Hanoi, Vietnam, in 2007, the M.Sc. degree from the Asian Institute of Technology, Thailand, in 2015, and the Ph.D. degree from the Department of Civil and Environmental Engineering, Sejong University, South Korea, in 2020. He is currently a Postdoctoral Researcher with Sejong University Academy Cooperation Foundation, Sejong University. He is also a Lecturer with the Faculty of Water Resources Engineering, Thuyloi University. His current research interests include machine learning applications in hydrological problems, radar-based rainfall analysis, urban hydrology, and water supply and drainage systems.

XUAN HIEN LE received the Ph.D. degree from the Department of Disaster Prevention and Environmental Engineering, Kyungpook National University (KNU), South Korea, in 2020. He is currently a Postdoctoral Researcher with the Emergency Management Institute, KNU. He is also a Lecturer with the Faculty of Water Resources Engineering, Thuyloi University, Vietnam. He is interested in the fields of hydrology, landslides, water supply, and drainage systems. His current research interest includes exploitation and application of deep learning techniques in solving hydrological problems.

JAE-YEONG HEO is currently pursuing the Ph.D. degree with the Department of Civil and Environmental Engineering, Sejong University. His current research interests include machine learning and data analysis applications in hydrological problems, rainfall-runoff modeling, and urban hydrology.

DEG-HYO BAE received the B.Sc. degree from Yonsei University, South Korea, in 1983, and M.Sc. and Ph.D. degrees from the University of Iowa, USA, in 1989 and 1992, respectively. From 1992 to 1994, he was a Researcher with the U.S. Department of Agriculture-ARS, USA. From 1994 to 1996, he was a Senior Researcher with Yonsei Industrial Technology University, South Korea. From 1996 to 2001, he worked as an Associate Professor with Changwon National University, Changwon, South Korea. He is currently a Professor with the Department of Civil and Environmental Engineering, Sejong University, South Korea. He is also the President of the Korean Water Resources Association. He has published many articles in international journals and been involved in many research projects in South Korea. His research interests include atmospheric and surface runoff interaction studies, climate change and variability studies, real-time flood and flash flood forecasting system development, radar-based rainfall analysis, hydrologic applications, GIS-based water resource engineering, and machine learning applications.