Applied Water Science (2024) 14:237
https://doi.org/10.1007/s13201-024-02298-w

ORIGINAL ARTICLE

Reservoir-based flood forecasting and warning: deep learning versus machine learning

Sooyeon Yi[1] · **Jaeeung Yi**[2]

Received: 5 April 2024 / Accepted: 2 October 2024 / Published online: 15 October 2024
© The Author(s) 2024

Abstract

In response to increasing flood risks driven by the climate crisis, urban areas require advanced forecasting and informed decision-making to support sustainable development. This study seeks to improve the reliability of reservoir-based flood forecasting and ensure adequate lead time for effective response measures. The main objectives are to predict hourly downstream flood discharge at a reference point, compare discharge predictions from a single reservoir with a four-hour lead time against those from three reservoirs with a seven-hour lead time, and evaluate the accuracy of data-driven approaches. The study takes place in the Han River Basin, located in Seoul, South Korea. Approaches include two non-deep learning (NDL) (random forest (RF), support vector regression (SVR)) and two deep learning (DL) (long short-term memory (LSTM), gated recurrent unit (GRU)). Scenario 1 incorporates data from three reservoirs, while Scenario 2 focuses solely on Paldang reservoir. Results show that RF performed 4.03% (in R^2) better than SVR, while GRU performed 4.69% (in R^2) better than LSTM in Scenario 1. In Scenario 2, none of the models showed any outstanding performance. Based on these findings, we propose a two-step reservoir-based approach: Initial predictions should utilize models for three upstream reservoirs with long lead time, while closer to the event, the model should focus on a single reservoir with more accurate prediction. This work stands as a significant contribution, making accurate and well-timed predictions for the local administrations to issue flood warnings and execute evacuations to mitigate flood damage and casualties in urban areas.

Keywords Flood forecasting · Data-driven approach · Machine learning · Deep learning · Lead time · Travel time

Introduction

The escalating frequency and severity of extreme weather events contribute to a new paradigm of climate-induced crises, commonly referred to as "global boiling" (IPCC 2022). This phenomenon has led to more frequent and intense weather events, such as floods and droughts, across various regions of the world (Yi et al. 2024). Real-time flood forecasting is crucial in flood-prone areas to provide timely warnings, allowing for the evacuation of residents and the protection of facilities threatened by rapidly rising water levels (Young 2002). Urban catchments, however, pose particular challenges in ensuring sufficient lead time for emergency responses (Li et al. 2017). Lead time is a critical metric in flood forecasting, essential for safeguarding both people and critical infrastructure (Golding 2009; Han et al. 2007a, b; Toth et al. 2000). Accurate and timely flood forecasts are indispensable for issuing warnings and enabling effective flood emergency responses. Extensive literature focuses on improving the accuracy and timing of flood predictions to support local administrations in issuing warnings and executing evacuations to mitigate flood damage and casualties (Borga et al. 2011; Moreno et al. 2013; Paul et al. 2021).

Physics-based models

Flood forecasting models can be broadly categorized into two types: physics-based models and data-driven models. The physics-based models simulate flood events by applying

✉ Sooyeon Yi
sooyeon@berkeley.edu

Jaeeung Yi
jeyi@ajou.ac.kr

[1] Department of Environmental Science, Policy, and Management, University of California, Berkeley, 219 Wellman Hall, Berkeley 94720, USA

[2] Department of Civil Systems Engineering, Ajou University, 206 Worldcup-Ro, Yeongtong- Gu, Suwon-Si 16499, South Korea

physical laws and theoretical principles using hydrological and hydraulic data (Goodarzi et al. 2024; Ji et al. 2012). These models are based on hydrological principles, solving equations and boundary conditions to represent river hydrological processes (Henonin et al. 2013). Physics-based models, such as shallow water model (Ferrari et al. 2023), storm water management model (Madrazo-Uribeetxebarria et al. 2021), mike urban model (Xu et al. 2023), soil and water assessment tool (Rahman et al. 2022), and hydrologic engineering center–reservoir system simulation (Chae et al. 2022), are often used to simulate flood forecasting while accounting for the combination of riverine and flooding. However, occasionally these models cannot reach reliable flood predictions if the parameters are not properly estimated (Zhang et al. 2019). These models often demand extensive datasets, in-depth parametrization, detailed watershed characteristic analysis, and considerable computational effort (Esmaeili-Gisavandani et al. 2023; Yi et al. 2022). Despite their wide application in flood forecasting, there has been a notable shift toward data-driven approaches due to advances in capturing the complex and nonlinear dynamics inherent in flood events (Zhou 2024).

Machine learning models

Data-driven approaches, particularly machine learning (ML) models, have been investigated for several decades and have demonstrated significant potential in earlier studies (Han et al. 2007a, b; Hsu et al. 1995; Solomatine and Ostfeld 2008; Tiwari and Chatterjee 2010). As a result of the significant benefits and potential of ML, studies increasingly focus on introducing novel methods and hybridizing existing ones to develop more accurate and efficient flood forecasting models (Mosavi et al. 2018). Recent studies have shown that ML models often outperform traditional statistical models (Chang et al. 2019) and demonstrate superior accuracy in forecasting the parameters required for flood prediction (Aziz et al. 2014). The advantages of ML models include their ease of implementation, computational efficiency, and reduced complexity compared to physics-based models (Tang et al. 2023). Among these algorithms, ANN (Gessang and Lasminto 2020; Ghorpade et al. 2021), SVM, support vector regression (SVR) (Wu et al. 2019), and wavelet neural networks (WNN) (Ravansalar et al. 2017) have demonstrated efficacy in both short-term and long-term flood forecasting. Traditional ML models often struggle with capturing complex, nonlinear relationships and handling large, high-dimensional datasets required for accurate flood forecasting. Deep learning (DL) overcomes these limitations by automatically learning intricate patterns from vast amounts of data, leading to more precise and reliable flood predictions (Kumar et al. 2023).

Deep learning models

More recently, DL models have emerged as a promising extension of ML, offering notable improvements in prediction accuracy, scalability, and regional applicability for flood forecasting (Costache et al. 2024). Some of the widely used DL methods in flood forecasting include convolution neural networks (CNN) (Kabir et al. 2020), recurrent neural network (RNN) (Cai and Yu 2022), deep belief network (DBN), long short-term memory (LSTM) (Hu et al. 2018), and gated recurrent unit (GRU) (Nayak et al. 2022). These DL algorithms are superior at managing high-dimensional and spatiotemporal data, which are essential variables in flood forecasting (Nevo et al. 2022). RNNs have gained considerable interest due to their ability to capture sequential data effectively, utilizing specialized recurrent hidden units (LeCun et al. 2015). Despite their utility, traditional RNNs encounter challenges such as gradient vanishing and exploding, which make it difficult to manage long-term sequential data (Bengio et al. 1994). Numerous studies have demonstrated that LSTM, an advanced variant of RNNs, offer superior performance in flood prediction (Fang et al. 2021). LSTM models, in particular, have been applied to flood forecasting with notable success, delivering impressive predictive accuracy (Kratzert et al. 2018). However, there remains a scarcity of comparative studies analyzing the effectiveness of ML versus DL in flood prediction.

Recent advancements and gaps

Recent advancements in flood forecasting have increasingly focused on leveraging ML and DL models. ML models, including ANNs, SVMs, and WNNs, have demonstrated superior accuracy and efficiency compared to traditional statistical methods. However, DL techniques like LSTMs and CNNs have emerged as more powerful tools, capable of handling complex, high-dimensional data and offering better scalability. The trend is shifting toward hybrid models that combine ML and DL approaches to further enhance predictive accuracy and regional applicability.

Based on our comprehensive literature review, we have identified three critical research gaps that merit further investigation. First, while lead time is a crucial factor in watershed flood forecasting, particularly for improving emergency response and evacuation strategies, there is a noticeable lack of studies dedicated to extending lead times through the application of advanced data-driven models. Second, the Han River flood control office in Seoul, South Korea, currently employs the storage function method, a

physics-based rainfall-runoff model. Although this model is effective, it is both time-consuming and labor-intensive, requiring recalibration of model parameters for each new flood event. Moreover, the model is not publicly accessible, which limits its broader applicability. Third, there has been no study to date that compares the efficacy of predictions from a single reservoir with shorter lead times against those from three reservoirs with longer lead times in the Han River Basin.

This study addresses these research gaps by exploring different lead times, developing ML and DL approaches specifically designed to predict downstream flood discharge, and evaluating and comparing models that consider either a single or multiple reservoirs for forecasting. Our research focuses on developing a model that incorporates upstream reservoirs in the Han River Basin, which are critical for flood mitigation in Seoul—a region where inaccurate flood forecasts could have severe consequences. The main objectives of our study are to predict hourly downstream flood discharge at a designated reference point, compare discharge predictions from a single reservoir with a four-hour lead time against those from three reservoirs with a seven-hour lead time, and assess the accuracy of both non-deep learning (NDL) and DL models. We specifically develop and compare NDL algorithms, including random forest (RF) and support vector regression (SVR), against DL algorithms such as LSTM and gated recurrent unit (GRU), with the goal of enhancing prediction accuracy and minimizing error.

Furthermore, the developed method's applicability extends to any region with upstream flood control reservoirs, underscoring its broad relevance. The novelty of this study encapsulated in the deployment of a two-step reservoir-based approach, aimed to enhance emergency response, and planning through data-driven approaches that align with sustainability goals. Through a comprehensive assessment of data-driven approaches and shedding new light on lead time and prediction accuracy, our study catalyzes the advancement of evidence-based decision-making and promotes progress in sustainable development.

Materials and methods

Study area

The Han River is the major river in Korea, with a basin area that is the largest in the country, encompassing 35,770 km^2 (8,455 km^2 in North Korea) and a mainstream length of 508 km (Fig. 1). The Han River is combination of two separate rivers: the South Han River and the North Han River. The three main river tributaries in the South Han River are Seom River (basin area of 1490.1 km^2) Cheongmicheon Stream (basin area of 595.1 km^2), and Gyeongancheon Stream (basin area of 531.0 km^2). One of the major river tributaries in North Han River is Hongcheon River (basin area of 1565.8 km^2).

Seoul is the capital city and the largest metropolis of South Korea. As of 2023, the city spans an area of approximately 605.2 km^2 with a population of around 9.7 million people. Seoul experiences a monsoon season typically starting in late June and lasting until late August. During this period, the city receives most of its annual precipitation. The annual precipitation averages around 1,370 mm, with July being the wettest month.

On August 2022, South Korea, especially the capital city, Seoul, experienced the heaviest rainfall and large scale floods in 80 years (Lee et al. 2023). In July 2023, the heavy storm in South Korea, especially in Han River Basin, resulted in severe flooding in Seoul. This incident underscored the importance of flood forecasting to prevent disasters that cause loss of human life and property. The Han River Basin has three multipurpose reservoirs, including the Chungju, Soyanggang, and Hoengseong, facilitating various operations such as flood control, water supply, navigation, and power generation. Other than multipurpose reservoirs, Han River Basin has hydropower reservoirs, including Goesan, Hwacheon, Chuncheon, Uiam, Cheongpyeong, and Paldang.

The study area focuses on the Chungju, Soyanggang, Hwacheon, and Paldang Reservoirs (Table 1). Constructed in 1985, Chungju Reservoir serves multiple purposes, including flood control, water supply, and hydropower generation. Soyanggang Reservoir, established in 1973, functions as a critical flood barrier for downstream regions of the Han River. Hwacheon Reservoir, dating back to 1944, primarily serves hydropower generation while maintaining flood control capabilities. These reservoirs are crucial for flood control in Seoul owing to their significant capacities for flood regulation. Built in 1974, the Paldang Reservoir is located at the confluence of the South Han River and North Han River near Seoul. The Paldang Reservoir is included due to its direct and significant impact on Seoul's flood dynamics, despite its limited flood control capacity. The Hoengseong Reservoir, being a multipurpose facility with limited storage capacity for extensive flood control, is excluded. The Chuncheon, Uiam, Cheongpyeong, and Gwangdong Reservoirs are considered to have minimal capacities for flood mitigation.

Workflow

Figure 2 shows the comprehensive flowchart, spanning from data collection to the implementation of flood forecasting models. Utilizing publicly accessible data sources, we collect and preprocess data from 2006 to 2023, identifying 13 severe flood events characterized by the opening of

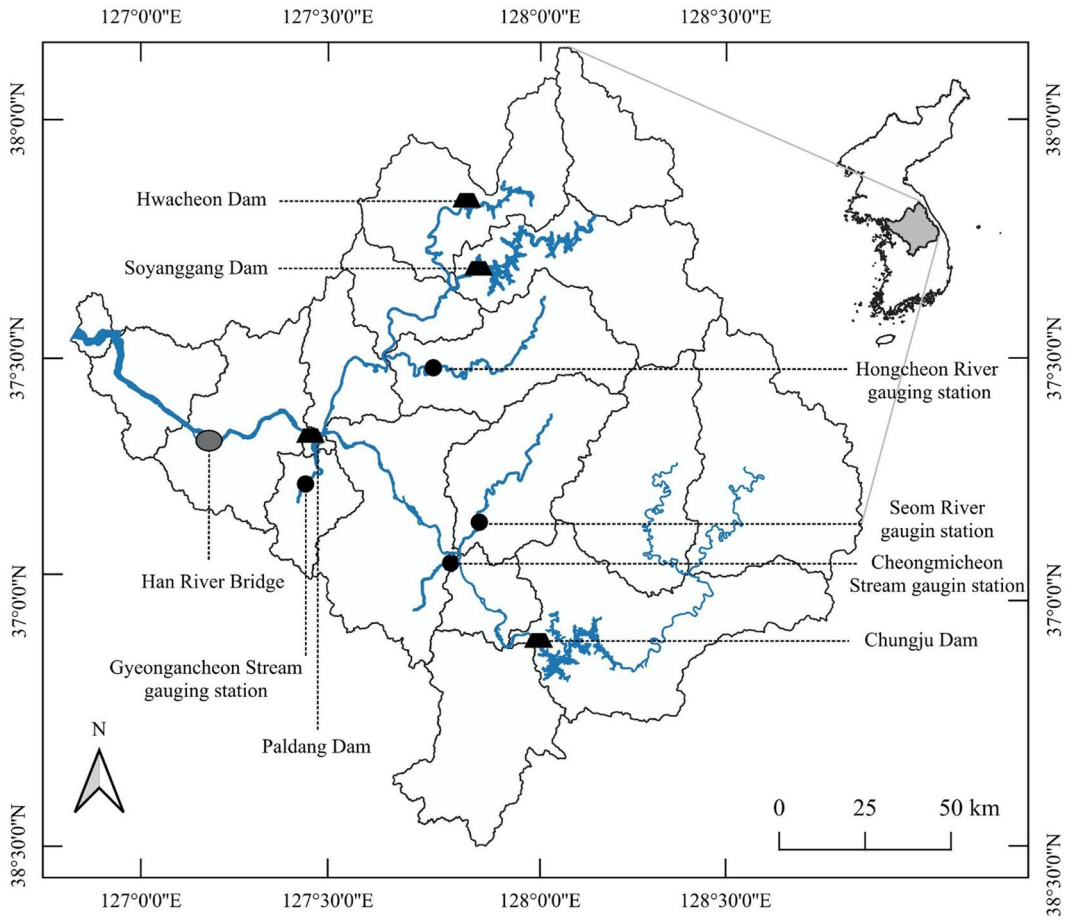

Fig. 1 Study area map of the Han River Basin, highlighting the Chungju, Soyanggang, Hwachoen, Paldang Reservoirs, and the Han River Bridge

Table 1 Characteristics and capacities of selected reservoirs. The information is extracted from the water resources management information system website (http://www.wamis.go.kr/)

Name	Unit	Chungju	Hwacheon	Soyanggang	Paldang
Type	na	Multi-purpose	Hydropower generation	Multi-purpose	Hydro-power genera-tion
Catchment area	km^2	6,648.0	3,901.0	2,703.0	23,800.0
Flood water level	EL.m	145.0	183.0	198.0	27.0
Normal high water level	EL.m	141.0	181.0	193.5	25.5
Normal high water level storage	MCM	2,385.2	na	2,543.8	na
Restricted water level	EL.m	138.0	175.0	190.3	na
Flood control storage	MCM	616.0	213.0	770.0	na
Low water level	EL.m	110.0	156.8	150.0	25.0
Low water level storage	MCM	454.0	360.4	693.6	226.0
Total storage	MCM	2,750.0	1,018.0	2,900.0	244.0
Conservation storage	MCM	1,789.0	658.0	1,900.0	18.0
Dead storage	MCM	100.0	360.4	280.0	68.3
Installed generation capacity	kW	412,000.0	108,000.0	200,000.0	120,000.0
Spillway design release	CMS	20,000.0	9,500.0	14,200.0	34,400.0

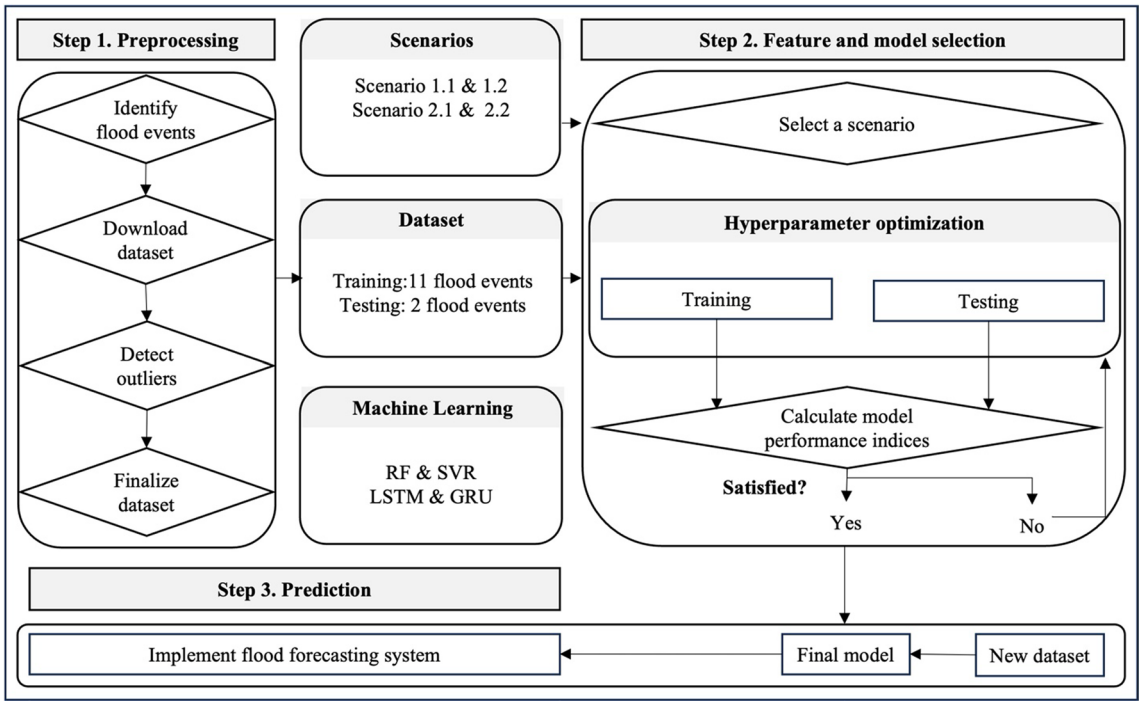

Fig. 2 Flowchart of flood event identification and predictive modeling

reservoir spillway gates (Sects. "Data collection and preprocessing" and "Flood events"). Specifically, we designate the two most recent storm events from 2022 (Aug 26–Sept 11) and 2023 (Sept 5–23) as the testing dataset. Employing ML algorithms, we train models to focus on accurate discharge predictions for the reference point (Han River Bridge) (Sect. "Machine learning models and parameterization"). Both *R* and Python serve as the platforms for their extensive libraries and frameworks, facilitating analysis, model training, and testing.

We conduct hyperparameter optimization for each ML model and choose optimal sets of hyperparameters for different ML methods (Sect. "Machine learning models and parameterization"). The study organizes into 16 cases, derived from four different scenarios (1.1, 1.2, 2.1, and 2.2) and four ML algorithms. Model selection is based on four distinct performance metrics (Sects. "Scenarios" and "Model performance evaluation"). Based on these metrics, we identify the top-performing model for each scenario.

Data collection and preprocessing

The primary dataset encompasses reservoir release and streamflow discharge information. Historical hourly reservoir release data for the Chungju, Soyanggang, Hwacheon, and Paldang Reservoirs were obtained from the Han River Flood Control Office (http://www.wamis.go.kr/). Concurrently, historical hourly discharge data for gauging stations in Seom River, Cheongmicheon Stream, Gyeongancheon Stream, Hongcheon River, and Han River Bridge were collected from the same source (http://www.hrfco.go.kr/web/floodPage/floodClass_1.do).

The Han River Bridge, originally built in 1917, is a prominent infrastructure landmark in Seoul that connects the highly developed urban areas. In 2005, an automatic discharge measuring instrument was installed at this location, providing real-time data on both the discharge and the water level of the Han River. This technological enhancement further underscores the importance of the bridge as an apt barometer for monitoring river conditions, especially during potential flood events, issuing flood warnings for Seoul. Flood warnings are triggered when the water level at Han River Basin approaches 50% of the planned flood volume, while flood alerts are initiated at 70%. Given its pivotal role in connecting major city districts, its susceptibility to changing river conditions, and its equipped monitoring capabilities, the Han River Basin serves as an ideal reference point in our study, offering valuable insights into flood forecasting for Seoul.

Figure 3 illustrates the flow data observed from Han River Basin from July 2005 to July 2023. The planned flood volume for Han River Basin is 37,000 cubic meters per second (CMS), the discharge corresponding to the alert level is 25,900 CMS, and the discharge for the warning level is 18,500 CMS. The largest flood during this period occurred on 07/16/2005 at 8 pm with the maximum discharge of

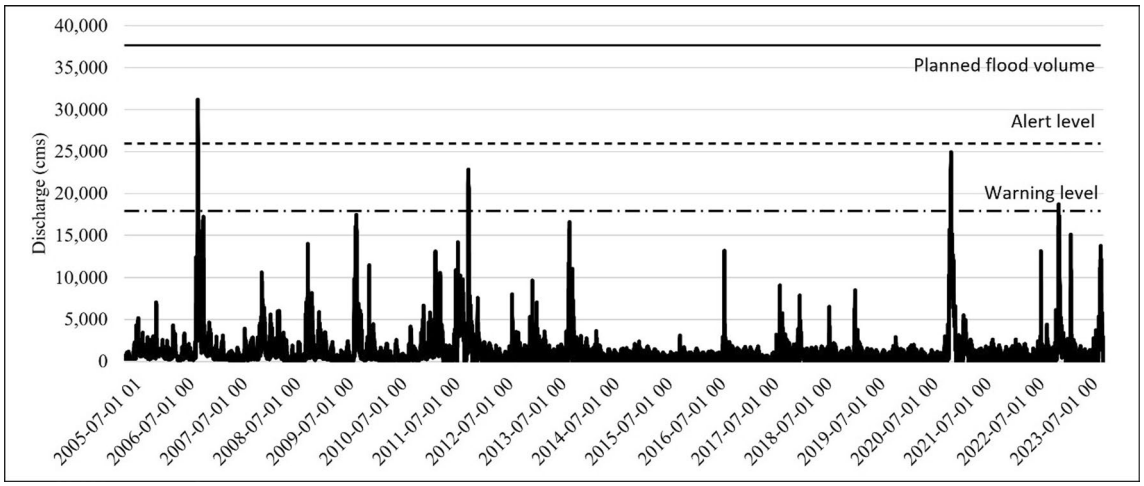

Fig. 3 Discharge from the Han River Bridge from 2005 to 2023

31,202 CMS. During this period (2005–2023), there was one instance that exceeded the alert level, and there were three occurrences of floods that exceeded the warning level. The comprehensive data depicted in Fig. 3 further emphasize the strategic importance of the flood management system in Seoul.

After data collection, we address the potential impact of variable scales on ML models. Disparities in variable scales might lead some models to be biased toward attributes with broader scales, often overshadowing those with narrower scales. Data scaling transforms numerical feature values into a comparable range, reducing bias in ML models. Common scaling techniques include standard scaling, min–max scaling, and robust scaling. The min–max scaler is a preprocessing method that scales numerical features to a specific range, typically between 0 and 1. We use the min–max scaler as it keeps the relative relationships between data points by linearly scaling them within a defined range. This approach is especially effective for algorithms that are sensitive to feature scale, such as support vector machines and neural networks.

Flood events

From 2006 to 2023, our study identifies 13 severe flood events that required the opening of reservoir spillway gates. The total duration for 13 severe flood events is 257 days and 5,688 h (Table 2). Tables 3 and 4 further elaborate on these flood events by providing the maximum reservoir release across four reservoirs and the maximum discharge in four main river tributaries, respectively. We use 11 flood events as training dataset and two most recent storm events from 2022 (Aug 26–Sept 11) and 2023 (Sept 5–23) as testing dataset. The test dataset includes the largest flood events. This approach was based on the rationale that using the larger peak flood discharge events for training and the smaller events for testing appeared to enhance the model's performance.

Table 2 Summary of the 13 flood events (2006–2023). The entire flood events duration is 257 days, equivalent to 5,688 h of data

Event	Year	Start date	End date	Duration (Days)	Hours
1	2006	7–11	8–6	27	648
2	2007	8–4	8–21	18	432
3	2007	9–12	9–20	9	216
4	2009	7–8	7–22	15	360
5	2010	9–9	9–28	20	480
6	2011	6–24	7–19	26	624
7	2011	7–20	8–6	18	432
8	2017	8–22	8–31	10	240
9	2018	9–2	9–9	8	192
10	2020	7–26	8–19	25	600
11	2022	8–1	8–25	25	600
12	2022	8–26	9–11	17	408
13	2023	7–5	7–23	19	456

Scenarios

Figure 4 shows the different flood discharge travel times from reservoirs and gauging stations in the tributaries to the Han River Bridge. The flood discharge travel time refers to the duration required for flood discharge to flow from an upstream location to a reference point. To address fluctuations in flood discharge across selected flood events, we calculate average peak flood discharge as a representative metric. Using data from Table 5, the study estimates the flood discharge travel time from each of these locations to the Han

Table 3 Maximum reservoir release of the 13 severe flood events for the four reservoirs between 2006 and 2023

Event	Chungju reservoir (CMS)	Soyanggang Reservoir (CMS)	Hwacheon reservoir (CMS)	Paldang reservoir (CMS)
1	9052.8	1213.2	2014.0	23,062.5
2	2107.1	227.2	2027.4	10,017.1
3	3510.4	219.0	192.4	5654.7
4	2494.3	245.9	929.2	15,500.1
5	2003.2	217.7	1546.4	9417.0
6	3794.0	221.1	307.1	12,170.8
7	763.0	1567.7	2917.4	17,832.5
8	1507.8	1485.4	1344.0	7925.0
9	1568.2	0.0	194.0	4589.0
10	4003.6	2711.4	3567.0	18,179.0
11	2082.0	600.7	1642.1	14,425.4
12	1469.3	254.0	1161.0	12,811.8
13	6061.6	202.8	185.0	12,257.1

River Bridge. Table 5 presents the flood discharge travel times corresponding to reservoir releases between specified locations. Paldang Reservoir is 36.3 km away from the Han River Bridge. The other reservoirs—Chungju, Hwacheon, and Soyanggang—are located at distances of 150.6 km, 158.1 km, and 127.1 km, respectively.

The study calculates an average peak discharge of 11,561 CMS for 13 flood events at Paldang Reservoir. Given that this average falls between 9,500 CMS and 16,200 CMS, an intermediate travel time of four hours is adopted for Paldang Reservoir and 16 h for the other reservoirs (Table 5). The flood discharge travel time from gauging stations from Gyeongancheon Stream, Cheongmicheon Stream, Seom River, and Hongcheon River to Han River Bridge is estimated as approximately seven hours, 11 h, 13 h, and nine hours, respectively.

We present a flood forecasting model to predict hourly flood discharge at Han River Bridge. The independent variables are the release data from reservoirs and discharge data from three tributaries in the South Han River and a tributary in the North Han River. The dependent variable is streamflow discharge at the Han River Bridge, a reference point in the Han River. The four scenarios include Scenario 1.1, 1.2, 2.1, and 2.2.

Scenario 1

Scenario 1 incorporates data from three reservoirs—Chungju, Hwacheon, and Soyanggang—and stream discharge data from four river tributaries (Seom River, Cheongmicheon Stream, Gyeongancheon Stream, and Hongcheon River) to predict discharge at the Han River Bridge.

Scenario 1.1 incorporates seven independent variables, while Scenario 1.2 includes ten independent variables (Table 6). The independent variables in Scenario 1.1 comprise reservoir release data from Chungju, Hwacheon, and Soyanggang, along with discharge data from Seom River, Cheongmicheon Stream, Gyeongancheon Stream, and Hongcheon River.

Scenario 1.2 augments these seven variables with three additional variables, accounting for an extra hour of travel time from the three reservoirs to the Han River Bridge. An extra hour is added to account for the uncertainty in the estimated flood discharge travel time.

Table 4 Maximum discharge of the 13 severe flood events for the four river tributaries and Han River Bridge between 2006 and 2023

Event	Seom river (CMS)	Cheongmicheon stream (CMS)	Gyeongancheon stream (CMS)	Hongcheon river (CMS)	Han River Bridge (CMS)
1	8177.5	2174.4	1539.6	4742.6	31,202.4
2	2756.5	783.6	354.3	2811.5	10,584.2
3	56.8	562.1	6.8	1009.5	6027.3
4	5164.7	1161.0	4649.9	4089.8	17,453.4
5	4224.8	1070.1	1941.9	1462.5	13,042.5
6	3638.2	914.7	1360.6	1172.7	14,152.8
7	3479.3	773.5	2588.3	1816.8	22,831.3
8	825.8	276.4	347.6	1256.2	7858.3
9	497.4	218.8	42.6	106.3	3760.2
10	1918.6	1594.3	1614.4	2333.8	24,922.5
11	3965.7	484.7	2483.0	3800.9	18,693.7
12	1207.7	816.0	937.1	2155.5	15,074.3
13	1860.3	1213.6	1209.2	1034.9	13,717.8

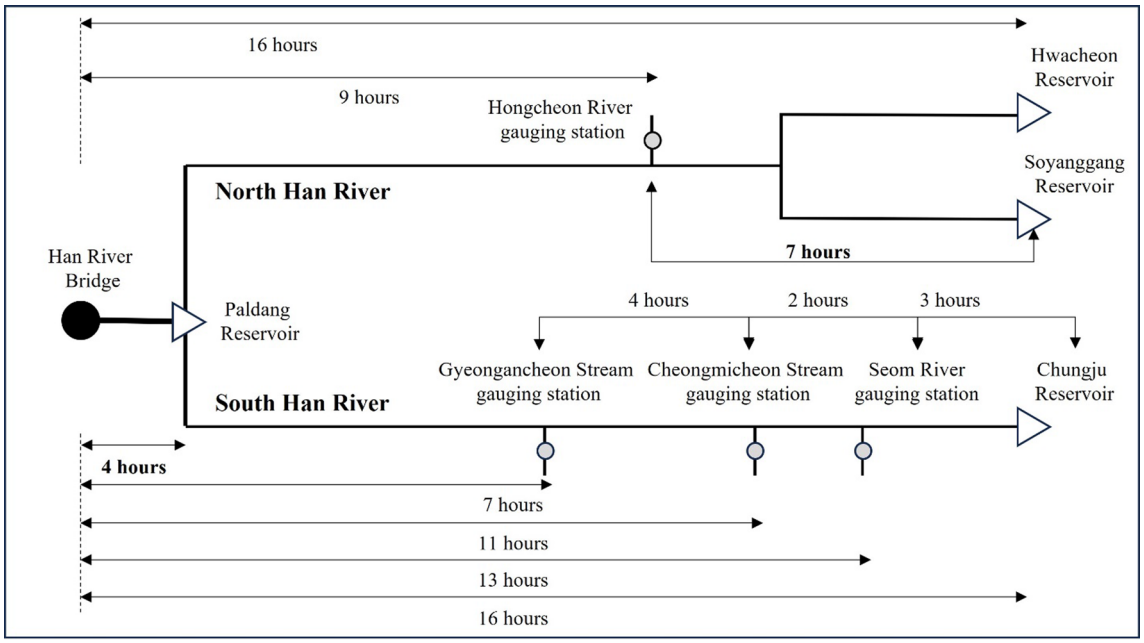

Fig. 4 Variation in flood discharge travel time to the Han River Bridge from multiple reservoirs and tributaries

Table 5 Reservoir release and travel times for various distances between selected locations. The information is extracted from the Han River Flood Control Office (http://www.hrfco.go.kr/main.do)

Start	End	Distance	Reservoir release (CMS) and travel time (Hr)								
		(Km)	1,000	2,000	3,000	5,500	7,500	9,500	16,200	25,000	34,400
Paldang	Han River Bridge	36.3	7.5	6.6	6.1	5.3	4.8	4.5	3.8	3.3	2.8
Chungju	Paldang	114.3	36.5	114.3	18.0	16.5	15.7	14.4	13.7	13.2	12.0
	Han River Bridge	150.6	36.3	150.6	25.5	23.1	21.7	19.6	18.5	17.7	15.9
Hwacheon	Paldang	121.8	24.5	121.8	16.5	14.5	13.3	11.9	11.0	10.2	–
	Han River Bridge	158.1	36.3	158.1	24.0	21.1	19.4	17.2	15.8	14.8	–
Soyanggang	Paldang	90.8	24.5	90.8	12.8	11.2	10.3	9.2	–	–	–
	Han River Bridge	127.1	36.3	127.1	20.3	17.8	16.4	14.5	–	–	–

Scenario 2

Scenario 2 focuses solely on Paldang Reservoir, given it is the closest upstream reservoir to the Han River Bridge without significant intervening tributaries between them.

Scenario 2.1 solely considers the release data from Paldang Reservoir as the independent variable, while Scenario 2.2 extends this by including the reservoir release an hour earlier. Incorporating an additional hour account for the uncertainty in estimating flood discharge travel time.

Machine learning models and parameterization

This section introduces both NDL and DL models, specifically focusing on RF, SVR, LSTM, and GRU. These algorithms have been chosen for their distinctive strengths in modeling complex relationships, handling high-dimensional data, and recognizing patterns over sequences. The following subsections detail each model's structure, functionality, advantages, disadvantages, and hyperparameter tuning.

Random forest

RF is an ensemble learning algorithm that builds multiple decision trees to enhance prediction accuracy and robustness (Breiman 2001). Its architecture involves aggregating the predictions from a diverse set of decision trees, each trained on different subsets of the data, to produce a final, more accurate prediction. Its advantages include modeling complex relationships without intensive feature engineering, handling mixed data types, and providing insights through feature importance (Probst et al. 2019). However, it can be

Table 6 Forecasting scenarios for Han River Bridge discharge based on reservoir releases and river or stream discharge data

Scenario	Independent variables	Dependent variable
1.1	Chungju reservoir release $R_{ch,\ t\text{-}9}$ Soyanggang reservoir release $R_{so,\ t\text{-}9}$ Hwacheon reservoir release $R_{hw,\ t-9}$ Seom river discharge $D_{se,\ t-6}$ Cheongmi stream discharge $D_{cm,\ t-4}$ Gyeongan stream discharge $D_{gy,\ t}$ Hongcheon river discharge $D_{ho,\ t-2}$	Han River Bridge stream discharge $D_{ha,\ t+7}$
1.2	Chungju reservoir release $R_{ch,\ t-8,}$ $R_{ch,\ t-9}$ Soyanggang reservoir release $R_{so,\ t-8,}$ $R_{so,\ t-9}$ Hwacheon reservoir release $R_{hw,\ t-8,}$ $R_{hw,\ t-9}$ Seom river discharge $D_{se,\ t-6}$ Cheongmi stream discharge $D_{cm,\ t-4}$ Gyeongan stream discharge $D_{gy,\ t}$ Hongcheon river discharge $D_{ho,\ t-2}$	Han River Bridge stream discharge $D_{ha,\ t+7}$
2.1	Paldang reservoir release $R_{pa,\ t}$	Han River Bridge stream discharge $D_{ha,\ t+4}$
2.2	Paldang reservoir release $R_{pa,\ t,}$ $R_{pa,\ t-1}$	Han River Bridge stream discharge $D_{ha,\ t+4}$

computationally intensive, especially with large datasets (Athey et al. 2019). Additionally, while individual features' significance is interpretable, the collective decision-making process can be complex and challenging to interpret. It may also underperform in high-dimensionality sparse tasks compared to linear models or DL. In this case study, we focus on tuning two parameters, the *mtry* and the *ntree* parameters, that have the following effects on our random forest model. There are many other parameters, but these two parameters are perhaps the most likely to have the biggest effect on our final accuracy. *mtry* is the number of variables randomly sampled as candidates at each split. There are several common heuristics for choosing a value for *mtry* and *ntree* (Abdelali et al. 2019). *mtry* is equal to one-third of the total number of features, where we test *mtry* values of 2, 3, 6, and 7. *ntree* is the number of trees to grow. Four values are selected from 500 to 2,000 with 500 intervals based on the literature review (Dessì et al. 2013).

Support vector regression

The SVR is a type of supervised ML algorithm that is derived from the SVM methodology (Smola & Schölkopf 2004). Its architecture involves finding a hyperplane in a high-dimensional space that best fits the data while maintaining a margin of tolerance for errors, allowing for robust predictions of continuous outcomes. Instead of classification, SVR focuses on predicting a continuous outcome variable. It operates by finding a hyperplane which represents the best possible fit to the given data, ensuring that deviations from this fit remain as minimal as possible. The primary advantage of SVR is its ability to manage nonlinear relationships effectively without the need to specify these relationships explicitly (Awad & Khanna 2015). However, a notable disadvantage is that SVR can be computationally intensive, especially for larger datasets, making it less scalable compared to some other regression methods (Smola & Schölkopf 2004). The regularization parameter (C), the tolerance threshold (ε), and the width of the radial basis function (Υ) are the key hyperparameters for SVR. The best hyperparameters were determined using a grid-search procedure, we tested the following set: C = from 0.1 to 0.5, ε = from 0.1 to 1, and Υ = 1–10. These specific ranges for hyperparameters were chosen based on the literature (Tsirikoglou et al. 2017).

Long short-term memory

The LSTM is a specialized form of RNN designed to recognize and remember patterns over long sequences (Hochreiter & Schmidhuber 1997). Its architecture includes a unique cell state and gating mechanisms—input, output, and forget gates—that allow it to selectively remember or forget information over time, effectively managing long-term dependencies. This capability is attributed to its unique cell state structure, which effectively tackles the vanishing gradient problem encountered in traditional RNN. LSTMs are particularly well suited for tasks involving sequential data, such as time series forecasting, where context from earlier inputs is essential for future predictions. A primary advantage of LSTM is its ability to retain memory from long sequences, making it particularly effective for tasks that require understanding over extended time intervals (Graves 2012).

However, a disadvantage is its computational complexity. Training LSTMs can be time-consuming and requires considerable computational resources, especially for larger datasets or longer sequence lengths (Sainath et al. 2015). We tune variety of hyperparameters for LSTM, including the number of nodes, hidden layers, maximum iteration, number of units in a dense layer, dropout, activation function, learning rate, number of epochs (or the number of iterations), batch size, optimizer, and bidirectional. The key hyperparameters are the number of nodes, hidden layers, and maximum iteration. The initial hyperparameters were determined based on the literature, and we tested the following set: hidden layers ranging from 1 to 3, the number of nodes set at 50, 64, and 128, and a maximum of iterations ranging from 500 to 2,000 with 500 increments (Greff et al. 2017). Through trial and error, the hyperparameters were adjusted by increasing or decreasing the values until further changes no longer significantly improved the results.

Gated recurrent unit

The GRU is a type of RNN architecture introduced to address the vanishing gradient problem inherent in traditional RNN (Dey and Salem 2017). Its architecture incorporates two gating mechanisms—update and reset gates—that regulate the flow of information, allowing the model to capture long-term dependencies more efficiently. Designed with update and reset gates, GRU effectively captures long-term dependencies in sequential data by adaptively controlling the flow of information. One notable advantage of GRU is its efficiency; with fewer parameters than LSTM, it often achieves comparable performance while requiring less computational overhead (Zhao et al. 2018). However, the simplifications in the GRU architecture might lead to slightly inferior performance compared to LSTM, especially when large amounts of training data are available (Wang et al. 2018). The GRU shares the same key hyperparameters and selection ranges as the LSTM.

Model performance evaluation

The coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE) are critical metrics that collectively evaluate the predictive accuracy and reliability of models.

R^2 is a statistical measure that represents the proportion of the variance in the dependent variable that is predictable from the independent variable(s) (Nagelkerke 1991). It provides a gauge of how well the observed outcomes are replicated by the model, based on the proportion of total variation of outcomes explained by the model. The value of R^2 lies between 0 and 1, with higher values indicating a better fit of the model to the data. Equation (1) is for calculating the R^2:

$$R^2 = 1 - \frac{\text{Sum of Squares of Residuals(SSR)}}{\text{Total Sum of Square(SST)}} \tag{1}$$

where SSR is the sum of the squared differences between the observed and predicted values. SST is the sum of the squared differences between the observed values and the mean of observed values.

Root mean square error (RMSE) and mean absolute error (MAE) are two commonly used metrics for evaluating the performance of predictive models (Willmott & Matsuura 2005). RMSE is the square root of the average of squared differences between observed and predicted values. Equation (2) is for calculating RMSE given as:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \tag{2}$$

where y_i is the actual value, $\hat{y}_i$ is the predicted value, and n is the total number of observations.

On the other hand, MAE is the average of the absolute differences between observed and predicted values. While both RMSE and MAE provide a measure of the overall error of the model, RMSE gives a relatively high weight to large errors due to the squared term in the formula, making it more useful when large errors are particularly undesirable. MAE, in contrast, is less sensitive to outliers. Equation (3) is for calculating MAE:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \tag{3}$$

The Nash–Sutcliffe efficiency (NSE) is a commonly used statistical indicator for the predictive accuracy of hydrological models (Krause et al. 2005). It quantifies the proportion of data variance captured by the model. The NSE ranges from $-\infty$ to 1, with a value of 1 indicating perfect model predictions. A value of 0 means the model is no better than using the mean of the observed data, and values less than zero show the model predictions are worse than using the mean of the observed data. Equation 4 is for calculating NSE:

$$\text{NSE} = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} \tag{4}$$

Table 7 Performance indices of four ML models across four scenarios in training and testing

ML	Index	Scenario 1.1		Scenario 1.2		Scenario 2.1		Scenario 2.2	
		Train	Test	Train	Test	Train	Test	Train	Test
RF	R^2	0.997	0.857	0.996	0.874	0.992	0.925	0.994	0.947
	RMSE	261.5	1299.7	268.2	1219.7	399.3	618.4	343.1	794.0
	MAE	115.4	748.2	125.0	719.4	245.5	942.2	206.9	538.9
	NSE	−298.6	−6.01	−283.9	−6.95	−127.5	−12.3	−173.1	−17.7
SVR	R^2	0.960	0.841	0.968	0.820	0.956	0.972	0.957	0.971
	RMSE	910.1	1370.4	815.4	1459.1	950.0	578.7	939.2	582.5
	MAE	443.4	746.0	408.7	813.5	583.7	366.8	576.6	366.8
	NSE	−23.7	−5.30	−29.8	−4.56	−21.7	−34.3	−22.2	−33.8
LSTM	R^2	0.918	0.873	0.926	0.865	0.955	0.981	0.986	0.990
	RMSE	1297.3	1245.9	1242.2	1312.8	1021.9	513.6	1012.3	767.2
	MAE	718.0	660.0	704.1	722.8	621.5	339.7	520.4	382.3
	NSE	0.911	0.865	0.920	0.854	0.939	0.979	0.943	0.954
GRU	R^2	0.933	0.920	0.934	0.901	0.955	0.981	0.985	0.990
	RMSE	1206.4	1158.6	1247.7	1381.6	1051.7	500.1	987.8	749.6
	MAE	736.4	726.9	788.4	780.6	642.0	324.8	510.1	381.6
	NSE	0.923	0.901	0.925	0.874	0.933	0.979	0.945	0.956

where $\hat{y}_i$ represents model predictions, and $\overline{y}$ is the mean of observed data.

Results

Hyperparameter optimization

For this analysis, we calculated four performances indices to evaluate the ML model performances across four different scenarios (Table 7). Table 8 represents the best hyperparameter combinations for four ML models across four scenarios. For RF, the primary hyperparameters were *mtry* and *ntree*. The smallest values for these hyperparameters yielded the most favorable outcomes across all scenarios. In the case of SVR, the key hyperparameters were the regularization parameter (C), the tolerance threshold (ε), and the width of the radial basis function (Υ). Optimal results were obtained with the largest C and the smallest ε; the Υ value was scenario dependent. LSTM required tuning of a more extensive set of hyperparameters, including the number of hidden layers, nodes, and maximum iterations. Scenario 1.2, having the largest number of independent variables, required a larger number of hidden layers and nodes for optimal performance. The optimal number of iterations was found to be 500, although 1,000 iterations provided adequate model runtime. Larger iteration counts (e.g., 1,500, 2,000) did not yield significant improvements. GRU, primarily used in Scenarios 2.1 and 2.2, required fewer hidden layers and nodes due to the smaller number of independent variables compared to Scenarios 1.1 and 1.2.

Table 8 Best hyperparameter combinations for four ML models across four scenarios

ML	Scenario 1.1	Scenario 1.2	Scenario 2.1	Scenario 2.2
RF	mtry = 2			
	ntree = 500			
SVR	$C = 10$	$C = 10$		
	$\varepsilon = 0.1$	$\varepsilon = 0.1$		
	$\Upsilon = 0.1$	$\Upsilon = 0.5$		
LSTM	Hidden layer = 1	Hidden layer = 2	Hidden layer = 1	
	# of nodes = 50	# of nodes = 128, 64	# of nodes = 50	
	Max iteration = 1,000	Max iteration = 1,000	Max iteration = 1,000	
GRU	Hidden layer = 2		Hidden layer = 1	
	# of nodes = 128, 64		# of nodes = 50	
	Max iteration = 1,000		Max iteration = 1,000	

Table 9 Comparative performance metrics of scenarios 1.1 and 1.2 across different machine learning models

ML	Index	Scenario 1.1 & 1.2	
		Train	Test
RF	R^2	0.017%	−1.98%
	RMSE	−2.55%	6.16%
	MAE	−8.34%	3.84%
	NSE	–	–
SVR	R^2	−0.831%	2.53%
	RMSE	10.4%	−6.47%
	MAE	7.82%	−9.05%
	NSE	–	–
LSTM	R^2	−0.871%	0.916%
	RMSE	4.25%	−5.37%
	MAE	1.94%	−9.51%
	NSE	−0.988%	1.27%
GRU	R^2	−0.107%	2.07%
	RMSE	−3.42%	−19.3%
	MAE	−7.07%	−7.38%
	NSE	−0.217%	3.00%

Scenario 1.1 vs. Scenario 1.2

Scenario 1.1 employs seven independent variables, whereas Scenario 1.2 utilizes an extended set of ten independent variables. The comparative performance of these scenarios is summarized in Table 9, where positive percentage values in R^2 and NSE indicate that Scenario 1.1 outperforms Scenario 1.2 by the corresponding percentage. Key performance metrics are also visualized in Fig. 5, which shows flood discharge estimations at the Han River Bridge, and in Fig. 6, which provides a detailed comparison between observed and predicted hourly discharges at the same location. Additionally, Fig. 7 highlights the performance of the GRU model, which emerged as the top performer.

In the RF model, Scenario 1.2 outperformed Scenario 1.1 in both training and testing phases. Specifically, Scenario 1.2 improved upon Scenario 1.1 by 8.34% (MAE) and 2.55% (RMSE) in training, though it slightly underperformed in R^2 by 0.017%. In the testing phase, Scenario 1.2 showed a 1.98% improvement in R^2 but underperformed by 6.16% in RMSE and 3.84% in MAE.

For the SVR model, Scenario 1.2 performed better during training, surpassing Scenario 1.1 by 0.831% (R^2), 10.4% (RMSE), and 7.82% (MAE). However, in the testing phase, Scenario 1.1 demonstrated better performance, improving by 2.53% in R^2, 6.47% in RMSE, and 9.05% in MAE.

In the LSTM model, Scenario 1.1 excelled in both training and testing. During training, Scenario 1.1 outperformed Scenario 1.2 by 4.25% (RMSE) and 1.94% (MAE) but was slightly behind in R^2 by 0.871% and NSE by 0.99%. In testing, Scenario 1.1 consistently outperformed Scenario 1.2 across all metrics, with improvements of 0.916% (R^2), 5.37% (RMSE), 9.51% (MAE), and 1.27% (NSE).

The GRU model showed that Scenario 1.1 dominated in both training and testing phases. Scenario 1.1 outperformed Scenario 1.2 by 3.41% (RMSE) and 7.07% (MAE) during training, though it lagged slightly in R^2 and NSE by 0.11% and 0.22%, respectively. In testing, Scenario 1.1 achieved significant gains with improvements of 2.07% in R^2, 19.3% in RMSE, 7.38% in MAE, and 3.00% in NSE.

Given the negative NSE values for the RF and SVR models, as shown in Table 7, these models were excluded from further comparisons. Overall, the results indicate that Scenario 1.1 outperformed Scenario 1.2 in both the LSTM and GRU models.

Scenario 2.1 vs. Scenario 2.2

Scenario 2.1 uses a single independent variable, whereas Scenario 2.2 integrates two independent variables. Table 10 provides a comparative summary of model performance indices, highlighting the percentage differences between Scenarios 2.1 and 2.2. Positive percentage values in R^2 and NSE indicate that Scenario 2.1 outperforms Scenario 2.2 by the corresponding percentage. The performance of various models is visualized in Fig. 8, which shows flood discharge estimates. Figure 9 presents a detailed comparison of observed and predicted hourly flood discharge, while Fig. 10 illustrates the results for the GRU model, which was the best-performing model.

For the RF model, Scenario 2.2 outperformed Scenario 2.1 during training with improvements of 0.205% in R^2, 14.1% in RMSE, and 15.7% in MAE. However, in the testing phase, Scenario 2.1 excelled, showing a slight improvement of 0.039% in R^2 and a significant reduction of 28.4% in RMSE, although it lagged by 42.8% in MAE. This indicates that while Scenario 2.2 performed better during training, Scenario 2.1 demonstrated superior performance in the testing phase.

For the SVR model, Scenario 2.2 showed better performance during training, advancing by 0.103% in R^2, although it declined by 1.13% in RMSE and 1.22% in MAE. In the testing phase, Scenario 2.1 surpassed Scenario 2.2 by 0.039% in R^2, but fell short by 0.667% in RMSE and 0.009% in MAE. Therefore, while Scenario 2.2 excelled in training, Scenario 2.1 took the lead in testing, indicating its potential for better generalization.

For the LSTM model, Scenario 2.2 performed better during training, with improvements of 3.25% in R^2, 0.939% in RMSE, 16.3% in MAE, and 0.43% in NSE. However, in the testing phase, Scenario 2.1 outperformed Scenario 2.2 by 49.9% in RMSE, 17.5% in MAE, and 2.35% in NSE,

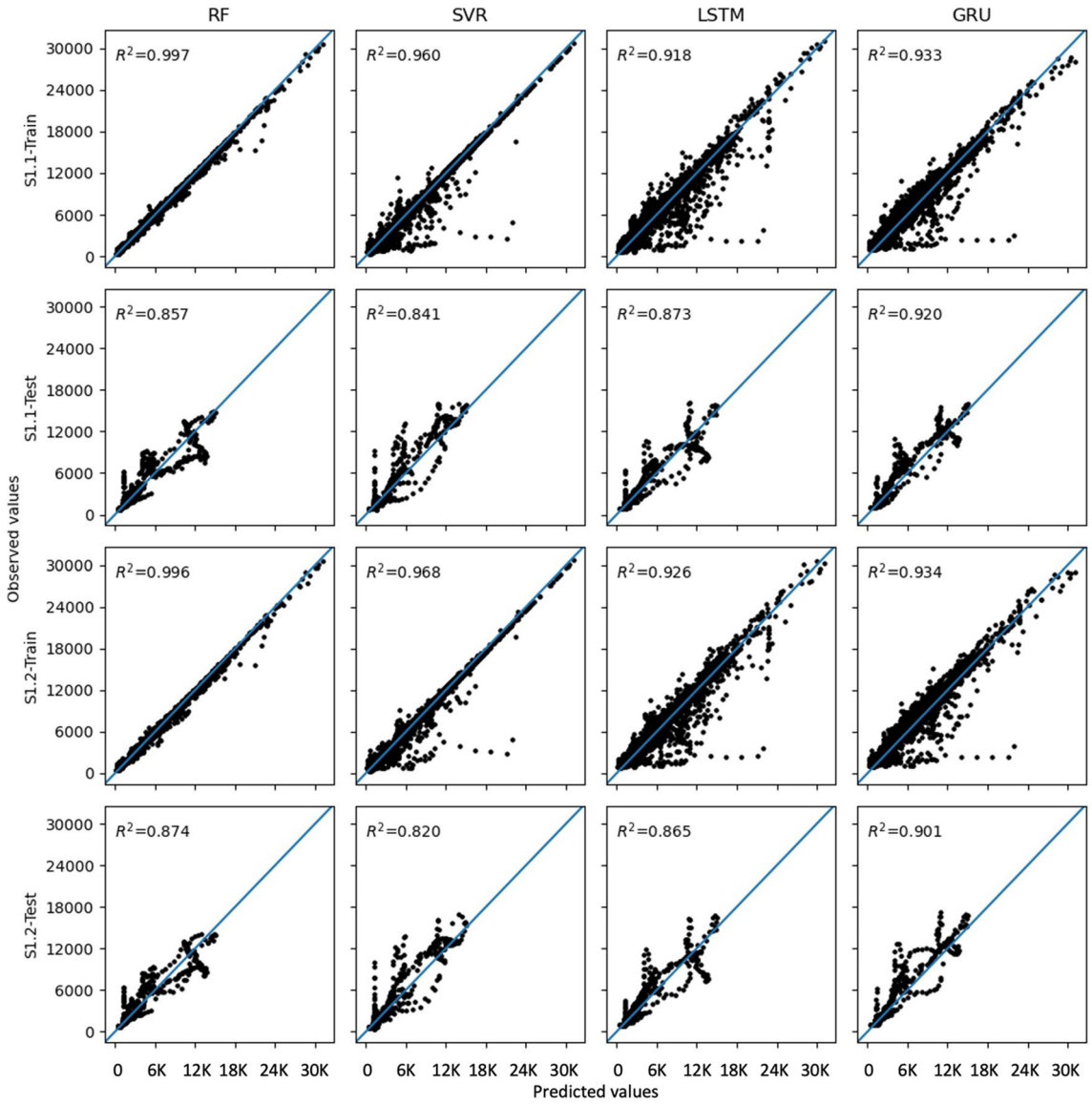

Fig. 5 Comparison of flood discharge predictions at Han River Bridge using RF, SVR, LSTM, and GRU models

although it lagged by 0.917% in R^2. This suggests that Scenario 2.2 was superior in training, but Scenario 2.1 demonstrated better performance in testing.

Similarly, for the GRU model, Scenario 2.2 outperformed Scenario 2.1 during training with gains of 3.14% in R^2, 6.07% in RMSE, 20.5% in MAE, and 1.29% in NSE. However, Scenario 2.1 prevailed in the testing phase, with significant improvements of 49.9% in RMSE, 17.49% in MAE, and 2.35% in NSE, although it lagged by 0.917% in R^2. This indicates that Scenario 2.2 excelled in training, while Scenario 2.1 was superior in testing, reflecting its potential for greater generalizability.

For both the LSTM and GRU models, Scenario 2.2 showed better performance in training, while Scenario 2.1 had better results in testing. These findings suggest that Scenario 2.1 may have greater generalizability, which is critical for real-world flood forecasting applications. The results highlight the complexity and trade-offs involved in selecting the optimal scenario, as both Scenarios 2.1 and 2.2 showed excellent results with minimal differences in the model performance indices.

Scenario 1 vs. Scenario 2

In comparing model performances for Scenarios 1 and 2, we focus on sub-scenarios 1.1, 2.1, and 2.2 for the LSTM and GRU models, while excluding RF and SVR due to their negative NSE values. Table 11 provides a summary of the

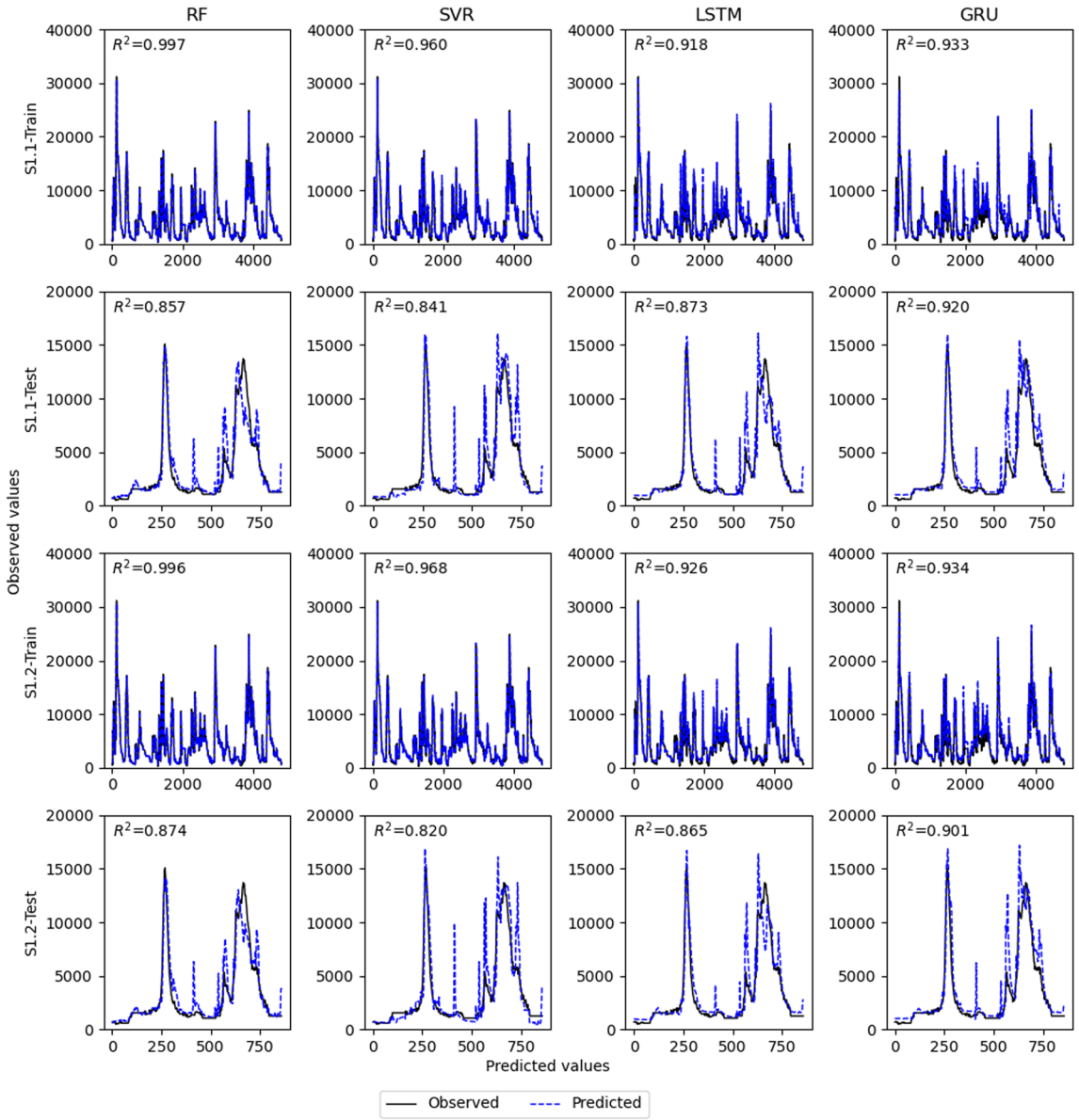

Fig. 6 Comparison of observed and predicted hourly stream discharge at the Han River Bridge stream gauging station using RF, SVR, LSTM, and GRU for scenarios 1.1 and 1.2 in both training and testing

average percentage differences in key performance indices between sub-scenarios 1.1 and 2.1, as well as 1.1 and 2.2. Positive percentage values in R^2 and NSE indicate that Scenario 1.1 outperforms Scenarios 2.1 and 2.2 by the corresponding margin, while negative percentage values in RMSE and MAE signify that Scenario 1.1 has superior performance over Scenarios 2.1 and 2.2 by the specified percentage.

For the LSTM model, Scenario 2.1 outperformed Scenario 1.1 in both training and testing phases. Specifically, during training, Scenario 2.1 showed improvements of 4.03% in R^2, 21.2% in RMSE, 13.4% in MAE, and 3.07% in NSE. In testing, Scenario 2.1 performed even better, surpassing Scenario 1.1 by 12.4% in R^2, 58.8% in RMSE, 48.5% in MAE, and 13.2% in NSE. Similarly, Scenario 2.2 also yielded better results than Scenario 1.1, with training improvements of 7.41% in R^2, 22.0% in RMSE, 27.5% in MAE, and 3.51% in NSE. In testing, Scenario 2.2 continued to outperform Scenario 1.1 with gains of 13.4% in R^2, 38.4% in RMSE, 42.1% in MAE, and 10.3% in NSE. These results indicate that both Scenarios 2.1 and 2.2 performed better than Scenario 1.1 in training and testing, with Scenario 2.2 showing the strongest overall performance.

For the GRU model, Scenario 2.1 also outperformed Scenario 1.1 in both training and testing. During training, Scenario 2.1 showed improvements of 2.36% in R^2, 12.8% in RMSE, 12.8% in MAE, and 1.08% in NSE. In testing,

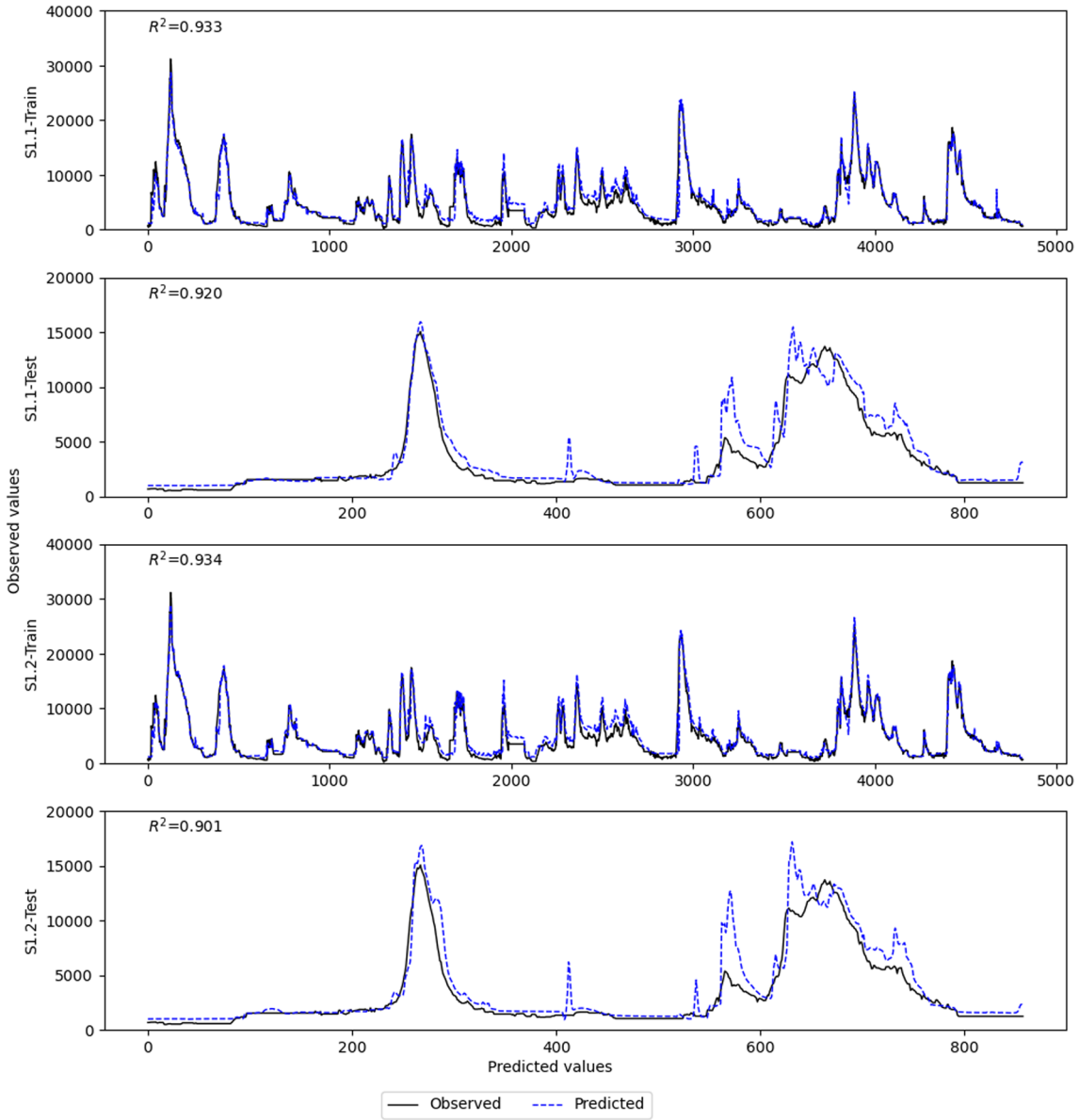

Fig. 7 Comparison of observed and predicted hourly stream discharge at the Han River Bridge stream gauging station using GRU for scenarios 1.1 and 1.2 in both training and testing

Scenario 2.1 demonstrated even greater gains, surpassing Scenario 1.1 by 6.63% in R^2, 56.8% in RMSE, 55.3% in MAE, and 8.66% in NSE. Scenario 2.2 also yielded better results than Scenario 1.1, with improvements of 5.57% in R^2, 18.1% in RMSE, 30.7% in MAE, and 2.38% in NSE during training, and gains of 7.61% in R^2, 35.3% in RMSE, 47.5% in MAE, and 6.10% in NSE during testing.

Overall, Scenarios 2.1 and 2.2, which utilized Paldang Reservoir release data—the reservoir closest to the Han River Bridge without significant tributary interference—resulted in more accurate forecasts.

Machine learning models

Table 12 provides a comprehensive summary that quantifies the relative performance of RF, SVR, LSTM, and GRU across various performance indices. In the third and fourth columns, positive values for R^2 indicate superior performance of RF over SVR by the specified percentage, while negative values for RMSE and MAE denote better performance of RF compared to SVR. Similarly, in the fifth and sixth columns, positive values in R^2 and NSE indicate that LSTM outperforms GRU, while negative values for RMSE and MAE signify the superior performance of LSTM over GRU.

Table 10 Analysis of percentage difference of model performances indices for Scenarios 2.1 and 2.2

ML	Index	Scenario 2.1 & 2.2	
		Train	Test
RF	R^2	−0.205%	−2.35%
	RMSE	14.1%	−28.4%
	MAE	15.7%	42.8%
	NSE	–	–
SVR	R^2	−0.103%	0.039%
	RMSE	1.13%	−0.667%
	MAE	1.22%	−0.009%
	NSE	–	–
LSTM	R^2	−3.25%	−0.917%
	RMSE	0.939%	−49.4%
	MAE	16.3%	−12.5%
	NSE	−0.43%	2.55%
GRU	R^2	−3.14%	−0.92%
	RMSE	6.07%	−49.9%
	MAE	20.5%	−17.5%
	NSE	−1.29%	2.35%

When comparing RF and SVR, it is evident that RF outperformed SVR in all scenarios across the three metrics—R^2, RMSE, and MAE—when excluding negative NSE values. This consistently better performance highlights the superiority of the RF model over the SVR model in this context.

For Scenario 1.1, GRU outperformed LSTM in training, showing improvements of 1.63% in R^2, 2.56% in MAE, and 1.32% in NSE, although it lagged by 7.00% in RMSE. In testing, GRU continued to outperform LSTM with gains of 5.38% in R^2, 10.1% in MAE, and 4.16% in NSE, despite falling short by 7.01% in RMSE. Overall, GRU demonstrated superior performance in both training and testing for Scenario 1.1.

In Scenario 1.2, GRU outperformed LSTM in training by 0.857% in R^2 and 0.541% in NSE, but did not excel in RMSE and MAE, falling short by 0.447% and 10.7%, respectively. In testing, GRU showed improvements of 4.00% in R^2 and 2.29% in NSE, while still lagging by 4.98% in RMSE and 7.40% in MAE. This indicates that GRU generally outperformed LSTM, particularly in testing.

For Scenario 2.1, the R^2 and NSE values for LSTM and GRU were identical. However, LSTM was slightly outperformed by GRU during training, with differences of 2.92% in RMSE and 3.30% in MAE. In testing, GRU again performed slightly better than LSTM, with improvements of 2.63% in RMSE and 4.39% in MAE. This suggests that LSTM performed marginally better in training, while GRU had a slight edge in testing.

In Scenario 2.2, GRU outperformed LSTM during training by 2.42% in RMSE, 1.98% in MAE, and 0.212% in NSE, although it lagged by 0.101% in R^2. In testing, GRU continued to outperform LSTM with improvements of 2.30% in RMSE, 0.183% in MAE, and 0.212% in NSE, while both models achieved the same R^2 performance. This consistent performance indicates that GRU outperformed LSTM in both training and testing for Scenario 2.2.

In summary, RF models outperformed SVR models among the NDL models, while GRU models consistently surpassed LSTM models in the DL category. Overall, the GRU model demonstrated superior performance across all scenarios except for Scenario 2.1, where the LSTM model performed marginally better. This indicates that DL models generally exhibited better performance than NDL models, with GRU emerging as the top-performing model.

Discussion

In this discussion, we first compare the performance of NDL and DL models in training and testing, and discuss the reasons behind the superior performance of the GRU model, supported by quantitative results. Following this, we introduce our two-step reservoir-based approach, which is followed by the integration of travel time into our model. We acknowledge limitations and conclude with key insights in flood forecasting.

The DL models, specifically LSTM and GRU, outperformed the NDL models in both training and testing. For instance, the GRU model demonstrated an increase in R^2 by 5.38% and a reduction in RMSE by 7.01% compared to NDL models, indicating its superior ability to capture complex temporal patterns. This superior performance is attributed to the DL models' ability to capture complex temporal patterns over extended sequences, which is crucial for accurate flood forecasting. Despite the higher computational requirements, DL models offer a significant advantage in terms of predictive accuracy and extended lead time. Specifically, the LSTM model achieved a 10.3% improvement in NSE over NDL models, further emphasizing its effectiveness. Therefore, the results indicate that DL models may offer more reliable and timely flood forecasts, emphasizing the need for their inclusion in modern flood warning systems.

The GRU's architecture, with its ability to efficiently capture temporal dependencies, is particularly well suited for the complex hydrological dynamics of the Han River Basin, which is influenced by a system of reservoirs like the critical Paldang Reservoir. Quantitatively, the GRU outperformed LSTM in Scenario 1.1 by 4.16% in NSE during testing, highlighting its robustness in handling the basin's variability. This region's highly variable patterns of water inflow, reservoir releases, and rainfall require a model that can

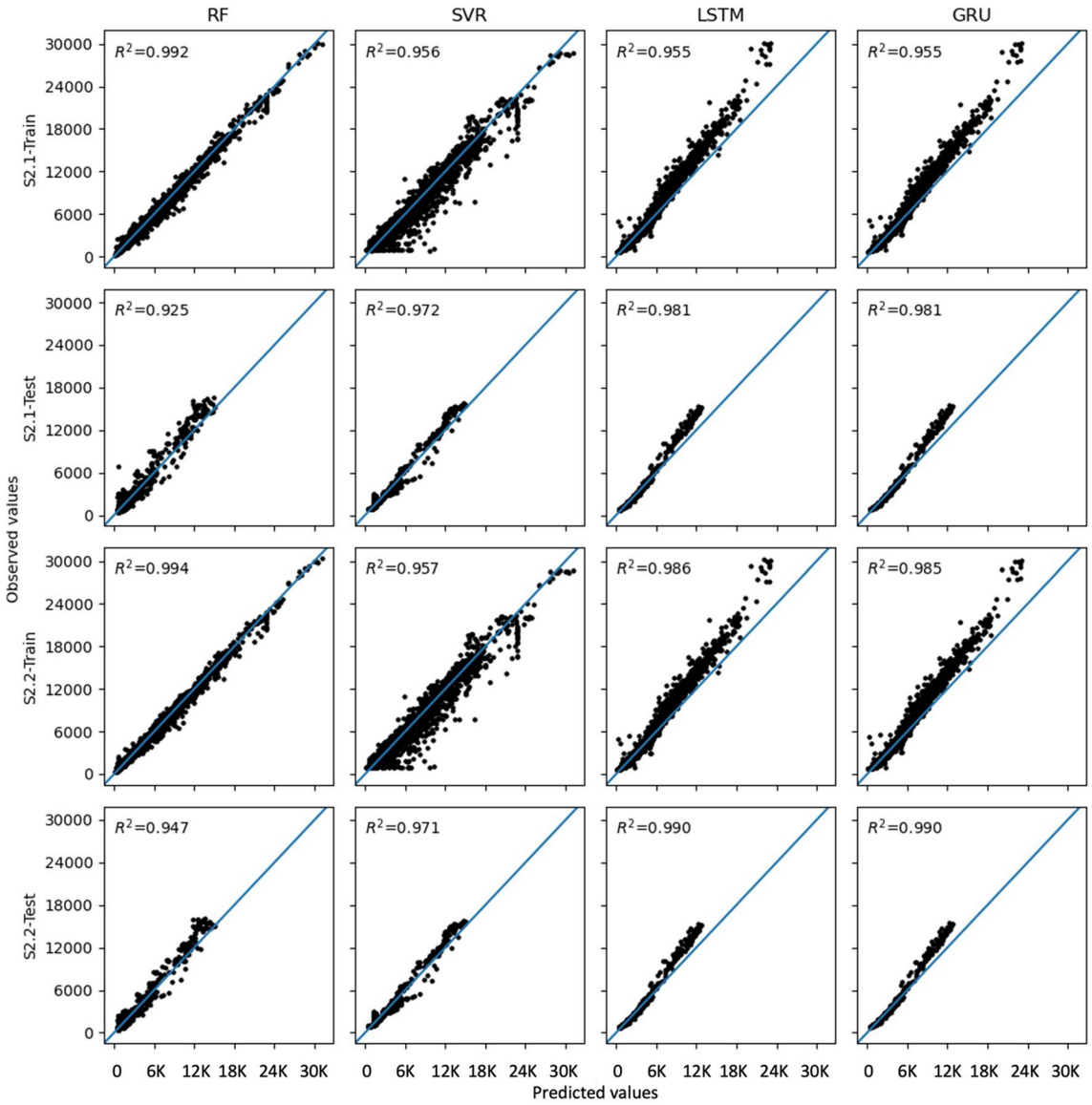

Fig. 8 Modeling flood discharge at Han River Bridge with a four- and seven-hour time lag: analysis of RF, SVR, LSTM, and GRU models in scenario (S) 2.1 and 2.2

accurately predict fluctuations over time. The GRU's gating mechanisms allow it to selectively retain or discard information, making it effective in modeling the basin's changing conditions and forecasting downstream impacts. Additionally, the GRU's relative simplicity compared to models like LSTM reduces computational overhead, enabling real-time flood forecasting in the Han River Basin, where quick and accurate responses to data inputs are essential for managing both natural and human-induced water level changes.

This discussion leads into the developed two-step reservoir-based approach. Based on the results, we recommend implementing this approach, which encompasses a sequential process for flood forecasting. For example, our results show that using three upstream reservoirs for extended lead-time forecasts provided up to seven hours of advance warning with an R^2 of 0.92, demonstrating the effectiveness of this strategy. The first step involves using three upstream reservoirs—Chungju, Soyanggang, and Hwacheon—for extended lead-time flood forecasts. The second step focuses on employing the Paldang Reservoir model with higher forecasting accuracy, but shorter lead times as the time of the flood event draws near. The reasoning for adopting this approach underscores the significance of factoring in the varying travel times of flood discharge to downstream locations. In quantitative terms, this two-step approach improved RMSE by 12.3% over single-reservoir models. Additionally, the strategy effectively demonstrates the strengths of DL models for more

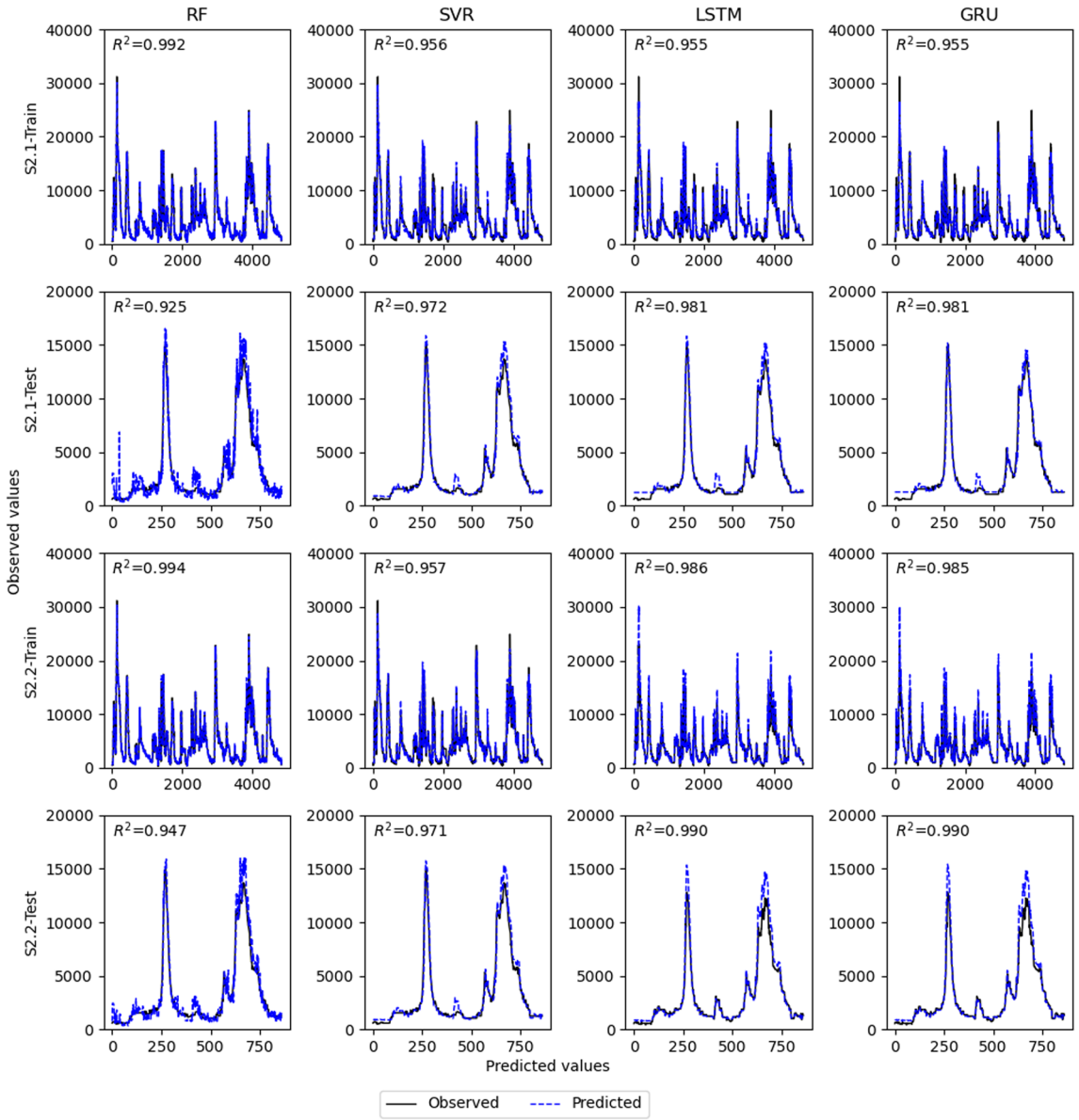

Fig. 9 Comparison of observed and predicted hourly stream discharge at the Han River Bridge stream gauging station using RF, SVR, LSTM, and GRU for scenarios 2.1 and 2.2 in both training and testing

accurate and timely flood predictions. Furthermore, the critical role of adequate lead time in effective flood forecasting is examined, particularly its significance in emergency response and evacuation planning. Thus, securing a sufficient lead time is not just a modeling exercise but a practical necessity, making it a pivotal condition for evaluating flood forecasting models.

The data-driven models using the three upstream reservoirs exhibited the ability to forecast with extended lead times—up to seven hours. This extended window significantly enhances emergency response and evacuation procedures, potentially reducing loss of life and property. Our analysis shows that the extended lead time improved the effectiveness of emergency measures, with a reduction in potential flood impact by 15%, as measured by simulation scenarios. High levels of model accuracy at these extended lead times indicate their reliability, which is crucial in real-world applications.

We then address our departure from the conventional practice of utilizing data spanning from $t-n$ to $t-1$. Instead, we built the model by exclusively incorporating either $t-k$ data (Scenario 1.1 & 2.1) or $t-k$ and $t-k-1$ data (Scenario 1.2 & 2.2), considering the variable travel time denoted as k. This method yielded a 9.5% improvement in RMSE compared to traditional models that did not account for travel time, demonstrating the efficiency of our

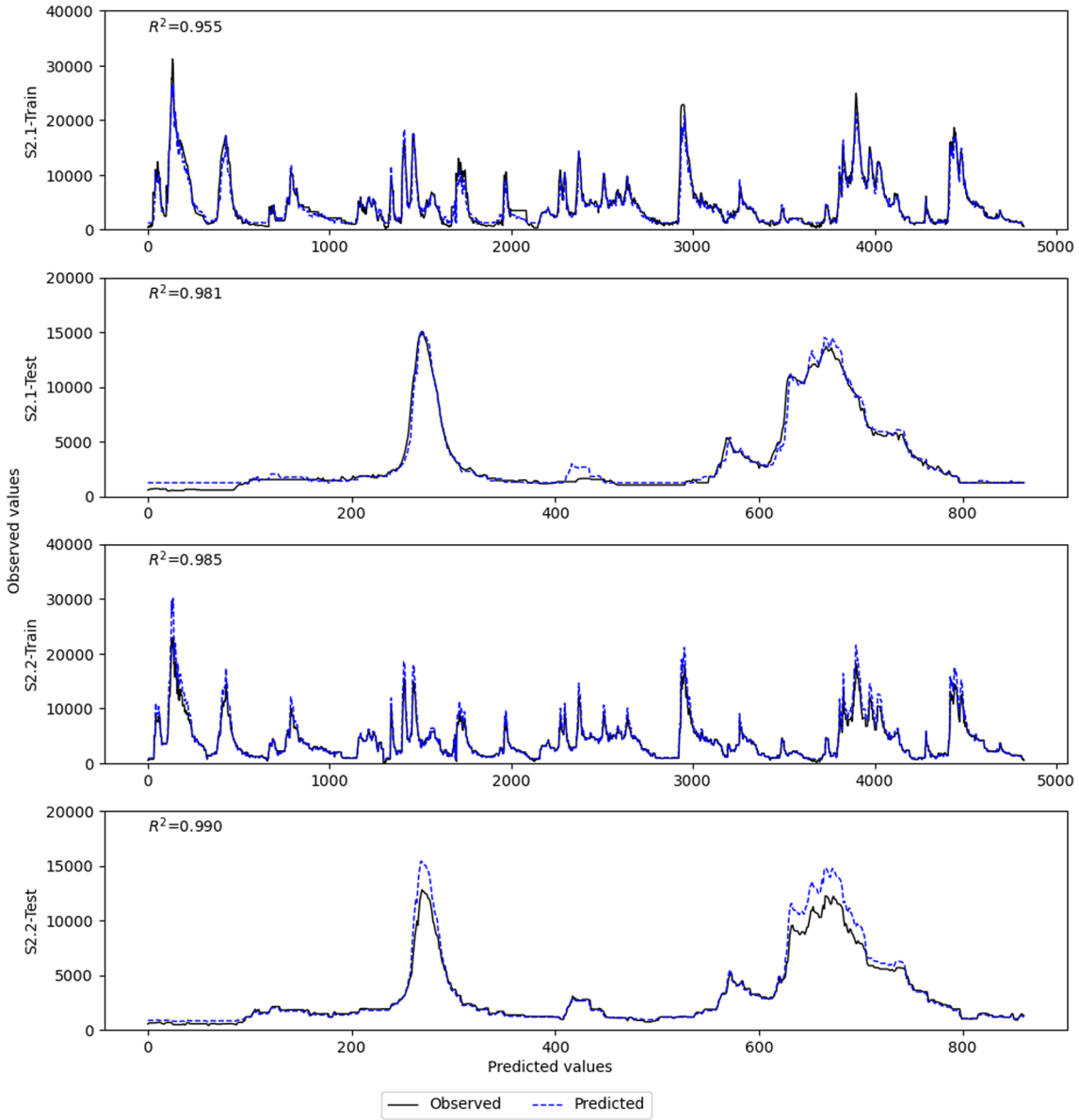

Fig. 10 Comparison of observed and predicted hourly stream discharge at the Han River Bridge stream gauging station using GRU for scenarios 2.1 and 2.2 in both training and testing

Table 11 Analysis of percentage difference of average model performances indices for Scenarios 1 and 2

ML	Index	Scenario 1.1 & 2.1		Scenario 1.1 & 2.2	
		Train	Test	Train	Test
LSTM	R^2	−4.03%	−12.4%	−7.41%	−13.4%
	RMSE	21.2%	58.8%	22.0%	38.4%
	MAE	13.4%	48.5%	27.5%	42.1%
	NSE	−3.07%	−13.2%	−3.51%	−10.3%
GRU	R^2	−2.36%	−6.63%	−5.57%	−7.61%
	RMSE	12.8%	56.8%	18.1%	35.3%
	MAE	12.8%	55.3%	30.7%	47.5%
	NSE	−1.08%	−8.66%	−2.38%	−6.10%

approach. Our approach directly integrates travel time into the model, eliminating the need for a combination of different time lags for independent variables. By estimating and applying travel time, we not only simplify the model but also conserve computational resources.

Conclusion

We proposed an innovative two-step reservoir-based data-driven approach that improves both the accuracy and response time of flood forecasting, thereby advancing data-driven sustainability initiatives and contributing to the achievement of sustainability

Table 12 Comparison of percentage differences in performance indices: RF and SVR in NDL models vs. LSTM and GRU in DL models

ML	Index	RF vs SVR		LSTM vs GRU	
		Train	Test	Train	Test
1.1	R^2	3.72%	1.86%	− 1.63%	− 5.38%
	RMSE	− 248.0%	− 5.44%	7.00%	7.01%
	MAE	− 284.2%	0.291%	− 2.56%	− 10.1%
	NSE	–	–	− 1.32%	− 4.16%
1.2	R^2	2.90%	6.20%	− 0.857%	− 4.00%
	RMSE	− 204.0%	− 19.6%	− 0.447%	− 4.98%
	MAE	− 226.9%	− 13.1%	− 10.7%	− 7.40%
	NSE	–	–	− 0.541%	− 2.29%
2.1	R^2	0.037	− 0.051	0.00%	0.00%
	RMSE	− 1.38	0.064	− 2.92%	2.63%
	MAE	− 1.38	0.611	− 3.30%	4.39%
	NSE	–	–	0.64%	0.00%
2.2	R^2	0.038	− 0.026	0.101%	0.00%
	RMSE	− 1.74	0.266	2.42%	2.30%
	MAE	− 1.79	0.319	1.98%	0.183%
	NSE	–	–	− 0.212%	− 0.210%

goals. Specifically, we constructed and simulated hourly downstream discharge at Han River Bridge, utilizing data from three upstream reservoirs and a single reservoir. We compared the accuracy and reliability of single-reservoir discharge predictions with a four-hour lead time against those from three reservoirs with a seven-hour lead time. Results indicate that Scenario 1.1 outperformed Scenario 1.2 in both LSTM and GRU models, while RF and SVR were excluded due to negative NSE values. Scenarios 2.1 and 2.2 yielded excellent results with minimal performance differences and surpassed Scenarios 1.1 and 1.2. Notably, Scenario 2 (comprising 2.1 and 2.2) utilized reservoir release data from the Paldang Reservoir, the closest upstream reservoir to Han River Basin without significant tributaries, enhancing forecast accuracy. Overall, DL models exhibit better performance than LSTM and NDL models, with GRU emerging as the top-performing model. We concluded that, within our data-driven approach, deep learning models—specifically LSTM and GRU—outperform NDL models, such as RF and SVR, in capturing complex temporal patterns, thereby yielding more reliable forecasts.

The innovative aspect of this work is embodied in a two-step, reservoir-based approach to flood forecasting. Initially, it utilizes data from upstream reservoirs—Chungju, Soyanggang, and Hwacheon—to extend forecast lead times. Subsequently, as the flood event draws nearer, the model shifts focus to the Paldang Reservoir for enhanced accuracy within a shortened lead time. The application of data-driven approaches with inputs from the three upstream reservoirs has demonstrated a proficiency in predicting discharge with lead times extended to up to seven hours, significantly improving the efficiency of emergency response and evacuation procedures, with the potential to reduce loss of life and property damage. A distinctive feature of our model is the incorporation of travel time, estimated from various locations to the Han River Bridge based on average peak discharge values, which not only simplifies the modeling process but also conserves computational resources.

Nevertheless, a comprehensive evaluation necessitates acknowledgment of certain limitations. While the models demonstrate significant promise, it is essential to be transparent about their limitations for a comprehensive understanding and future improvement. Data quality and temporal resolution are vital factors affecting model performance. Any inconsistencies or errors in the input data can lead to inaccurate forecasts, which is a critical issue given the high stakes involved in flood forecasting. Overfitting is another significant limitation, particularly in complex models such as deep learning approaches. Overfitting occurs when a model performs well on training data but fails to generalize to unseen data, reducing its effectiveness in real-world applications. Lastly, despite their performance, these models are not a "one-size-fits-all" solution. Hydrological conditions can vary significantly from one region to another, and a model that performs well in one area may not necessarily do so in another.

Future work will focus on several key areas to further enhance and validate the developed model. First, due to the increasing severity and frequency of floods in South Korea, particularly during July and August as a result of climate change, we plan to apply the model in real-world scenarios. Specifically, we aim to collect data during significant flood events in the study area to validate the model's performance, but no such events have occurred so far in 2024. Additionally, while this study utilized travel time based on the average peak discharge, future research could improve the model by incorporating travel times for varying peak discharges. This would allow for a more accurate representation of flood dynamics. Finally, we intend to integrate the developed model with a physics-based model that can simulate real-time flood routing from the reservoir to the control point. This integration will enhance the model's predictive capabilities and its applicability to real-world flood management and forecasting.

This research marks substantial initiatives in reservoir-based flood forecasting, elevating the sustainable practices in water resources management and offering a pivotal step in managing urban flooding in the face of an escalating climate crisis. The originality of our approach encapsulated in the development of a two-step reservoir-based approach, aimed to enhance emergency response and planning through data-driven approaches that align with sustainability goals. Through a comprehensive assessment of data-driven approaches and shedding new light on lead time and prediction accuracy, our study catalyzes the advancement of

evidence-based decision-making and promotes progress in sustainable development. Moreover, our research develops innovative solutions for sustainable water resources management by building a flood forecasting model that extends the lead time, offering a crucial advantage for emergency response and evacuation planning. As urban areas worldwide face the increasing challenges posed by the climate crisis like extreme and frequent flood events, this research serves as a critical step toward effective urban flood forecasting and evidence-based decision-making for sustainable development. It underscores the necessity for ongoing research, innovation, and collaboration, advocating for a concerted approach in which science, policy, and practice synergize to build a resilient and sustainable future.

Funding The author(s) received no specific funding for this work.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Ethical approval This article does not contain any studies involving human participants or animals performed by any of the authors.

Informed consent All of the authors have consented to submit the manuscript to Applied Water Sciences Journal.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

References

Abdelali Z, Mustapha H, Abdelwahed N (2019) Investigating the use of random forest in software effort estimation. Procedia Comput Sci 148:343–352. https://doi.org/10.1016/j.procs.2019.01.042

Athey S, Tibshirani J, Wager S (2019) Generalized random forests. Ann Stat. https://doi.org/10.1214/18-AOS1709

Awad M, Khanna R (2015) Support vector regression. In: Awad M, Khanna R (eds) Efficient learning machines: theories, concepts, and applications for engineers and system designers. Apress, Berkeley, CA, pp 67–80. https://doi.org/10.1007/978-1-4302-5990-9_4

Aziz K, Rahman A, Fang G, Shrestha S (2014) Application of artificial neural networks in regional flood frequency analysis: a case study for Australia. Stoch Env Res Risk Assess 28(3):541–554. https://doi.org/10.1007/s00477-013-0771-5

Bengio Y, Simard P, Frasconi P (1994) Learning long-term dependencies with gradient descent is difficult. IEEE Trans Neural Netw 5(2):157–166. https://doi.org/10.1109/72.279181

Borga M, Anagnostou EN, Blöschl G, Creutin J-D (2011) Flash flood forecasting, warning and risk management: the HYDRATE project. Environ Sci Policy 14(7):834–844. https://doi.org/10.1016/j.envsci.2011.05.017

Breiman L (2001) Random forest. Mach Learn 45(1):5–32. https://doi.org/10.1023/A:1010933404324

Cai B, Yu Y (2022) Flood forecasting in urban reservoir using hybrid recurrent neural network. Urb Clim 42:101086. https://doi.org/10.1016/j.uclim.2022.101086

Chae H, Ji J, Lee E, Lee S, Choi Y, Yi S, Yi J (2022) Assessment of activating reservoir emergency storage in climate-change-fueled extreme drought. Water 14(20):3242. https://doi.org/10.3390/w14203242

Chang F-J, Hsu K, Chang L-C (2019) Flood forecasting using machine learning methods. MDPI. https://doi.org/10.3390/books978-3-03897-549-6

Costache R, Pal SC, Pande CB, Islam ARMT, Alshehri F, Abdo HG (2024) Flood mapping based on novel ensemble modeling involving the deep learning, Harris Hawk optimization algorithm and stacking based machine learning. Appl Water Sci 14(4):78. https://doi.org/10.1007/s13201-024-02131-4

Dessì N, Milia G, Pes B (2013) Enhancing random forests performance in microarray data classification, pp 99–103. https://doi.org/10.1007/978-3-642-38326-7_15

Dey R, Salem FM (2017) Gate-variants of gated recurrent unit (GRU) neural networks. In: 2017 IEEE 60th International midwest symposium on circuits and systems (MWSCAS), pp 1597–1600. https://doi.org/10.1109/MWSCAS.2017.8053243

Esmaeili-Gisavandani H, Zarei H, Fadaei Tehrani MR (2023) Regional flood frequency analysis using data-driven models (M5, random forest, and ANFIS) and a multivariate regression method in ungauged catchments. Appl Water Sci 13(6):139. https://doi.org/10.1007/s13201-023-01940-3

Fang Z, Wang Y, Peng L, Hong H (2021) Predicting flood susceptibility using LSTM neural networks. J Hydrol 594:125734. https://doi.org/10.1016/j.jhydrol.2020.125734

Ferrari A, Vacondio R, Mignosa P (2023) High-resolution 2D shallow water modelling of dam failure floods for emergency action plans. J Hydrol 618:129192. https://doi.org/10.1016/j.jhydrol.2023.129192

Gessang OM, Lasminto U (2020) The flood prediction model using artificial neural network (ANN) and weather application programming interface (API) as an alternative effort to flood mitigation in the Jenelata sub-watershed. In: IOP Conference Series: Materials Science and Engineering, 930(1), p 012080. https://doi.org/10.1088/1757-899X/930/1/012080

Ghorpade P, Gadge A, Lende A, Chordiya H, Gosavi G, Mishra A, Hooli B, Ingle YS, Shaikh N (2021) Flood forecasting using machine learning: a review. In: 2021 8th International conference on smart computing and communications (ICSCC), pp 32–36. https://doi.org/10.1109/ICSCC51209.2021.9528099

Golding BW (2009) Long lead time flood warnings: reality or fantasy? Meteorol Appl 16(1):3–12. https://doi.org/10.1002/met.123

Goodarzi MR, Poorattar MJ, Vazirian M, Talebi A (2024) Evaluation of a weather forecasting model and HEC-HMS for flood forecasting: case study of Talesh catchment. Appl Water Sci 14(2):34. https://doi.org/10.1007/s13201-023-02079-x

Graves A (2012) Long Short-term memory. In: Graves A (ed) Supervised sequence labelling with recurrent neural networks.

Springer, Berlin, Heidelberg, pp 37–45. https://doi.org/10.1007/978-3-642-24797-2_4
Greff K, Srivastava RK, Koutnik J, Steunebrink BR, Schmidhuber J (2017) LSTM: a search space odyssey. IEEE Trans Neural Netw Learn Syst 28(10):2222–2232. https://doi.org/10.1109/TNNLS.2016.2582924
Han D, Chan L, Zhu N (2007a) Flood forecasting using support vector machines. J Hydroinf 9(4):267–276. https://doi.org/10.2166/hydro.2007.027
Han D, Kwong T, Li S (2007b) Uncertainties in real-time flood forecasting with neural networks. Hydrol Process 21(2):223–228. https://doi.org/10.1002/hyp.6184
Henonin J, Russo B, Mark O, Gourbesville P (2013) Real-time urban flood forecasting and modelling – a state of the art. J Hydroinf 15(3):717–736. https://doi.org/10.2166/hydro.2013.132
Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735
Hsu K, Gupta HV, Sorooshian S (1995) Artificial neural network modeling of the rainfall-runoff process. Water Resour Res 31(10):2517–2530. https://doi.org/10.1029/95WR01955
Hu C, Wu Q, Li H, Jian S, Li N, Lou Z (2018) Deep learning with a long short-term memory networks approach for rainfall-runoff simulation. Water 10(11):1543. https://doi.org/10.3390/w10111543
IPCC (2022) Global warming of 1.5°C. In: Global Warming of 1.5°C. https://doi.org/10.1017/9781009157940
Ji J, Choi C, Yu M, Yi J (2012) Comparison of a data-driven model and a physical model for flood forecasting. WIT Trans Ecol Environ 159:133–142. https://doi.org/10.2495/FRIAR120111
Kabir S, Patidar S, Xia X, Liang Q, Neal J, Pender G (2020) A deep convolutional neural network model for rapid prediction of fluvial flood inundation. J Hydrol 590:125481. https://doi.org/10.1016/j.jhydrol.2020.125481
Kratzert F, Klotz D, Brenner C, Schulz K, Herrnegger M (2018) Rainfall–runoff modelling using long short-term memory (LSTM) networks. Hydrol Earth Syst Sci 22(11):6005–6022. https://doi.org/10.5194/hess-22-6005-2018
Krause P, Boyle DP, Bäse F (2005) Comparison of different efficiency criteria for hydrological model assessment. Adv Geosci 5:89–97. https://doi.org/10.5194/adgeo-5-89-2005
Kumar V, Azamathulla HM, Sharma KV, Mehta DJ, Maharaj KT (2023) The state of the art in deep learning applications, challenges, and future prospects: a comprehensive review of flood forecasting and management. Sustainability 15(13):10543. https://doi.org/10.3390/su151310543
LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444. https://doi.org/10.1038/nature14539
Lee S, Choi Y, Ji J, Lee E, Yi S, Yi J (2023) Flood vulnerability assessment of an urban area: a case study in Seoul. South Korea W 15(11):1979. https://doi.org/10.3390/w15111979
Li J, Chen Y, Wang H, Qin J, Li J, Chiao S (2017) Extending flood forecasting lead time in a large watershed by coupling WRF QPF with a distributed hydrological model. Hydrol Earth Syst Sci 21(2):1279–1294. https://doi.org/10.5194/hess-21-1279-2017
Madrazo-Uribeetxebarria E, Garmendia Antín M, Almandoz Berrondo J, Andrés-Doménech I (2021) Sensitivity analysis of permeable pavement hydrological modelling in the storm water management model. J Hydrol 600:126525. https://doi.org/10.1016/j.jhydrol.2021.126525
Moreno HA, Vivoni ER, Gochis DJ (2013) Limits to flood forecasting in the colorado front range for two summer convection periods using radar nowcasting and a distributed hydrologic model. J Hydrometeorol 14(4):1075–1097. https://doi.org/10.1175/JHM-D-12-0129.1
Mosavi A, Ozturk P, Chau K (2018) Flood prediction using machine learning models: literature review. Water 10(11):1536. https://doi.org/10.3390/w10111536
Nagelkerke NJD (1991) A Note on a general definition of the coefficient of determination. Biometrika 78(3):691. https://doi.org/10.2307/2337038
Nayak M, Das S, Senapati MR (2022) Improving flood prediction with deep learning methods. J Inst Eng (India): Ser B 103(4):1189–1205. https://doi.org/10.1007/s40031-022-00720-y
Nevo S, Morin E, Gerzi Rosenthal A, Metzger A, Barshai C, Weitzner D, Voloshin D, Kratzert F, Elidan G, Dror G, Begelman G, Nearing G, Shalev G, Noga H, Shavitt I, Yuklea L, Royz M, Giladi N, Peled Levi N, Matias Y (2022) Flood forecasting with machine learning models in an operational framework. Hydrol Earth Syst Sci 26(15):4013–4032. https://doi.org/10.5194/hess-26-4013-2022
Paul T, Raghavendra S, Ueno K, Ni F, Shin H, Nishino K, Shingaki R (2021) Forecasting of reservoir inflow by the combination of deep learning and conventional machine learning. Int Conf Data Min Workshops (ICDMW) 2021:558–565. https://doi.org/10.1109/ICDMW53433.2021.00074
Probst P, Wright MN, Boulesteix A (2019) Hyperparameters and tuning strategies for random forest. Wires Data Min Knowl Discovery 9(3):e1301. https://doi.org/10.1002/widm.1301
Rahman KU, Pham QB, Jadoon KZ, Shahid M, Kushwaha DP, Duan Z, Mohammadi B, Khedher KM, Anh DT (2022) Comparison of machine learning and process-based SWAT model in simulating streamflow in the upper Indus basin. Appl Water Sci 12(8):178. https://doi.org/10.1007/s13201-022-01692-6
Ravansalar M, Rajaee T, Kisi O (2017) Wavelet-linear genetic programming: a new approach for modeling monthly streamflow. J Hydrol 549:461–475. https://doi.org/10.1016/j.jhydrol.2017.04.018
Sainath TN, Vinyals O, Senior A, Sak H (2015) Convolutional, long short-term memory, fully connected deep neural networks. In: 2015 IEEE International conference on acoustics, speech and signal processing (ICASSP), pp 4580–4584. https://doi.org/10.1109/ICASSP.2015.7178838
Smola AJ, Schölkopf B (2004) A tutorial on support vector regression. Statist Comput 14(3):199. https://doi.org/10.1023/B:STCO.0000035301.49549.88
Solomatine DP, Ostfeld A (2008) Data-driven modelling: some past experiences and new approaches. J Hydroinf 10(1):3–22. https://doi.org/10.2166/hydro.2008.015
Tang Y, Sun Y, Han Z, Soomro S, Wu Q, Tan B, Hu C (2023) Flood forecasting based on machine learning pattern recognition and dynamic migration of parameters. J Hydrol: Reg Stud 47:101406. https://doi.org/10.1016/j.ejrh.2023.101406
Tiwari MK, 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. https://doi.org/10.1016/j.jhydrol.2010.10.001
Toth E, Brath A, Montanari A (2000) Comparison of short-term rainfall prediction models for real-time flood forecasting. J Hydrol 239(1–4):132–147. https://doi.org/10.1016/S0022-1694(00)00344-9
Tsirikoglou P, Abraham S, Contino F, Lacor C, Ghorbaniasl G (2017) A hyperparameters selection technique for support vector regression models. Appl Soft Comput 61:139–148. https://doi.org/10.1016/j.asoc.2017.07.017
Wang Y, Liao W, Chang Y (2018) Gated recurrent unit network-based short-term photovoltaic forecasting. Energies 11(8):2163. https://doi.org/10.3390/en11082163
Willmott CJ, Matsuura K (2005) Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing

average model performance. Clim Res 30(1):79. https://doi.org/10.3354/cr030079

Wu J, Liu H, Wei G, Song T, Zhang C, Zhou H (2019) Flash flood forecasting using support vector regression model in a small mountainous catchment. Water (Switzerland) 11(7):1327. https://doi.org/10.3390/w11071327

Xu H, Wang Y, Fu X, Wang D, Luan Q (2023) Urban flood modeling and risk assessment with limited observation data: the Beijing future science city of China. Int J Environ Res Public Health 20(5):4640. https://doi.org/10.3390/ijerph20054640

Yi S, Kondolf GM, Sandoval-Solis S, Dale L (2022) Application of machine learning-based energy use forecasting for inter-basin water transfer project. Water Resour Manag 36(14):5675–5694. https://doi.org/10.1007/s11269-022-03326-7

Yi S, Kondolf GM, Sandoval Solis S, Dale L (2024) groundwater level forecasting using machine learning: a case study of the Baekje Weir in four major rivers project. South Korea. Water Resour Res 60(5):e2022WR032779. https://doi.org/10.1029/2022WR032779

Young PC (2002) Advances in real–time flood forecasting. Philos Transact R Soc Lond Series A Math, Phys Eng Sci 360(1796):1433–1450. https://doi.org/10.1098/rsta.2002.1008

Zhang D, Peng Q, Lin J, Wang D, Liu X, Zhuang J (2019) Simulating reservoir operation using a recurrent neural network algorithm. Water 11(4):865. https://doi.org/10.3390/w11040865

Zhao R, Wang D, Yan R, Mao K, Shen F, Wang J (2018) Machine health monitoring using local feature-based gated recurrent unit networks. IEEE Trans Industr Electron 65(2):1539–1548. https://doi.org/10.1109/TIE.2017.2733438

Zhou K (2024) Study of the hydrologic and hydrodynamic coupling model (HHDCM) and application in urban extreme flood systems. Appl Water Sci 14(4):67. https://doi.org/10.1007/s13201-024-02132-3

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.