The Role of Event Identification in Translating Performance Assessment of Time-Series Urban Flood Forecasting

Farzad Piadeh [a], Kourosh Behzadian [a,1], Amir Alani [a]

[a] *School of Computing and Engineering, University of West London, St Mary's Rd, London, W5 5RF, UK*

Abstract

Today, urban flood forecasting becomes a hydrological hot topic due to urbanisation, population growth, the staggering rise in weather extremes and its significant consequences such as economic, social and infrastructural losses. To address this, multiple time-series urban flood forecasting models have been developed recently. These models inevitably require relatively long-range and continuous time-series data. However, this database usually is mixed by the large share of unnecessary data, particularly dry weather conditions in which there is no flood occurrence. In this situation, the conventional model performance is assessed based on the entire database and consequently is diluted by prediction results of unnecessary data in practice. To overcome this, this paper aims to propose a framework to identify the three different approaches including rainfall-based, water level -based and hybrid method. In each method, model performance is determined based on a precise portion of data entitled events to remove the effects of dry weather conditions. Events are defined as a part of the database which represents urban flooding. Differences in these methods are illustrated through a real case study of urban flood forecasting in Ruislip's drainage system. The nonlinear autoregressive network with exogenous model is used for predictions of 15-minute, 1-hour, 2-hour and 3-hour steps ahead. Furthermore, mean absolute error, root mean square error, coefficient of determination and Nash–Sutcliffe model efficiency coefficient are evaluated as performance assessment indicators. The results reveal the role of event identification in the performance assessment of these models. While the conventional method shows the best performance, indicators are expected to become worth in the rainfall-based, the water level-based and hybrid methods, respectively.

Keywords: Data classification; Event identification; Flood forecasting; Realistic performance assessment

1. Introduction

Today, urban flooding is going to be the most destructive natural disasters due to its economic, environmental, and human losses in cities (Kourtis and Tsihrintzis, 2021). To overcome this, urban flood forecasting models have been utilised viably as an effective non-structural approach and an early warning system to mitigate the adverse effects of flood occurrence and reduce the associated risks (Accarino *et al.*, 2021). Multiple methods have been developed in recent decade based on data-driven models in which artificial intelligence (AI) and specifically neural networks (NN) are the most popular ones (Zounemat-Kermani *et al.*, 2020). These models are highly contributed to the advancement of more accurate predictions through mimicking complex flood processes with less data required and more cost-effective solutions (Mosavi *et al.*, 2018).

The performance of these models is typically evaluated based on the comparison between the model outputs and the corresponding measurements in the test data that are not used for the model development (i.e. unseen data) through a number of key performance indicators (KPIs) (Daal *et al.*, 2017). While these models use relatively long-range and continuous sets of time-series data for training and validation processes (Thrysøe *et al.*, 2019; Piadeh *et al.*, 2022), the test data used to estimate KPIs should be carefully selected to avoid unnecessary data particularly

[1] Corresponding author. Tel.: +44 (0) 20 8231 2466
E-mail address: kourosh.behzadian@uwl.ac.uk

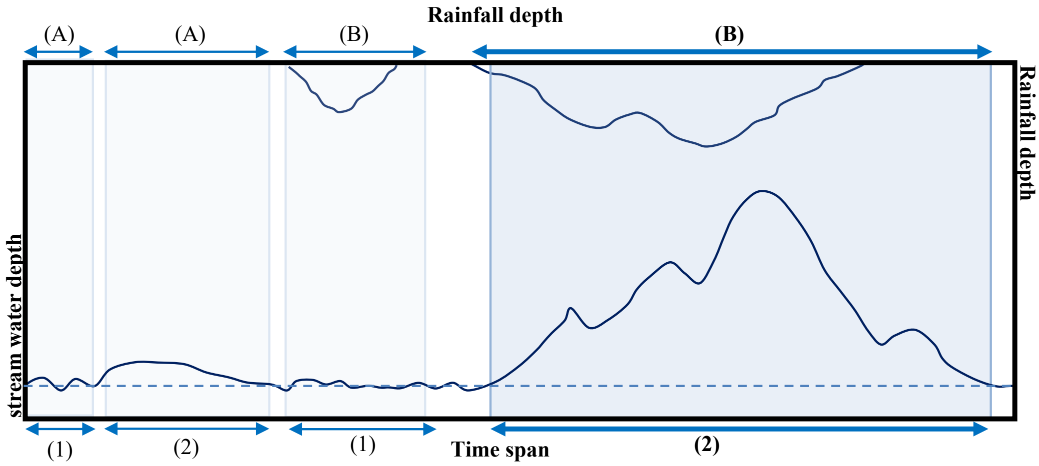
dry weather conditions with no flooding. However, the performance assessment has relied on the fact of including a long period of test data for KPIs that may contain unnecessary data. While some studies suggested the use of some KPIs such as false alarm ratio to provide screen methods for test data, there is lack of clear screening methods to distinguish between dry and wet weather data used to estimate KPIs.

There are a few approaches to identify flood events used for performance assessment of real-time models. While Darabi *et al.*, (2019) claimed that there is no inclusive guideline for flood event identification, DEFRA (2020) defined a flood event in a stream for the time when the water depth in the stream increases until it comes back to normal providing that rainfall occurs. Some works specifically used this time definition for flood even identification and estimated their model performance within the periods of flood events such as Adikari *et al.* (2021) which determined a three-class model performance defined by three ranges of annual peak runoff, runoff threshold for high flows, and runoff threshold for low flows. Lin *et al.* (2021) considered streamflow flood events based on rainfall duration which was then developed through adding rainfall depth by Huang *et al.* (2021). However, they failed to provide any details for screening or classification methods used to build their flood events. Hence, the present paper aims to propose a new framework for classification of rainfall runoff data to identify flood events used for performance assessment of the urban flood forecasting models. For this purpose, the proposed methodology is described in section 2 and its application is demonstrated and analysed through a real case study in the UK. Finally, the conclusions are drawn, and future studies are finally recommended.

2. Methodology for flood event identification and data classification methods

While a flood event in a stream is generally demonstrated by the time duration in which water depth rises and drops due of rainfall occurrence. The start and end points of this duration can determine the data boundaries of the flood event. Generally, rainfall intensity/depth and water depth at a stream/sewer chamber are the two main datasets used for data-driven flood forecasting models. The water depth can either increase or remain almost steady during wet or dry weather conditions. Rainfall depth can also be either zero or not during dry or wet weather conditions. Based on these assumptions for rainfall and water depth, a flood event can be recognised as one of the four states shown in Figure 1 as: (1) “Dry weather” (S1): no rainfall measured and trivial/no change in water depth, (2) “Upstream flow” (S2): no rainfall but water depth increase due to several reasons such as sanitary sewage discharged into combined sewerage, leakage/exfiltration or infiltration, (3) “No effective rainfall” (S3): rainfall occurs but it is evaporated, collected in ground depression/canopy before reaching the ground or infiltrated in the ground before converting to runoff in the catchment and discharging into sewer or stream, and hence there is no effective rainfall and runoff occur finally (4) “flood event” (S4): rainfall occurs and water depth rises and hence this is the only state for flood which needs to specify the start point of the flood event.

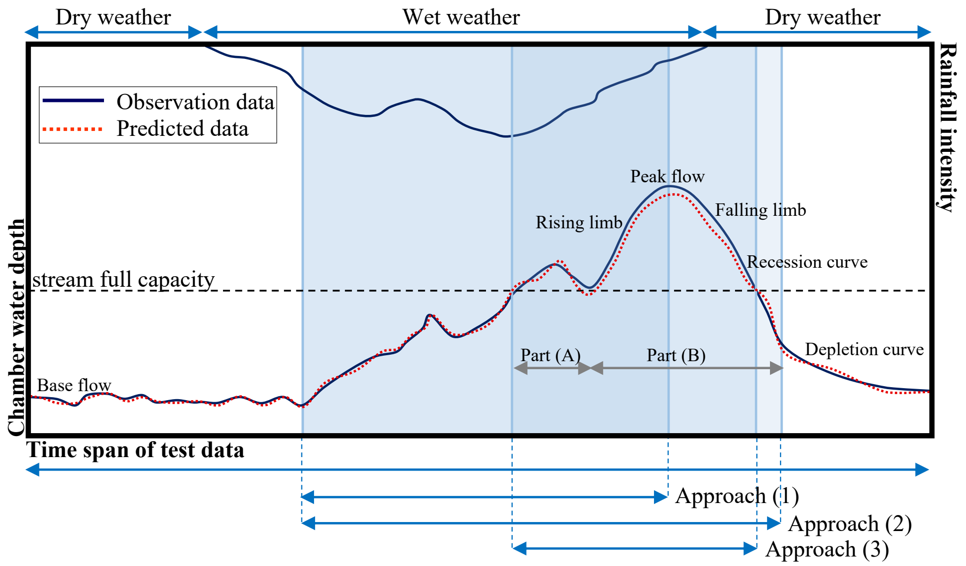
Three approaches defined here to identify the boundaries of a flood event are (1) conventional, (2) rainfall-based, (3) water level-based, and (4) alarm-based, as demonstrated in Figure 2 and Table 2. A conventional approach evaluates the model performance based on entire test data, whereas other approaches provide data classification methods to find flood events, remove unnecessary predicted data, and measure the only flood predicted data The flood boundaries in all introduced approaches are identified as (1) rainfall-based: it starts when rainfall contributes to the stream water level i.e. when water depth rises in the stream until it reaches to the flood peak, (2) water level-based approach: the boundary is between the time when water level rises as a result of rainfall and then returns to normal water depth level, and (3) Alarm-based approach: the boundaries are defined only when water level exceeds the full capacity of the stream. The rainfall approach shown in Algorithm 1 starts by capturing the first recorded rainfall intensity. Whenever rainfall intensity is recorded at zero value, the event is checked by the chamber's water depth rising to ensure that water level uprising occurrence is caused by rainfall. After confirmation, risen water depth data are captured until when the rising water depth is stopped, and water depth is positioned in a falling limb (See Figure 2).



Key:

State	Data (increase + and no change -)	
	Rainfall depth	Water depth
(S1): Dry weather – Unnecessary event	(A): -	(1): -
(S2): Upstream flow – Unnecessary event	(A): -	(2): +
(S3): No effective rainfall – Unnecessary event	(B): +	(1): -
(S4): Flood event	(B): +	(2): +

Figure 1. Schematic representation for classification of relationship between rainfall depth and stream's water depth data used for performance assessment of real-time flood forecasting models



Key:
Approach (1): Rainfall-based Approach (2): Runoff-based Approach (3): Alarm-based

Figure 2. Three data classification methods for flood event identification

Table 2: Characteristics of data classification approaches for flood event identification

Characteristics	Approach			
	Conventional	**Rainfall-based**	Water level **-based**	**Alarm-based**
Definition	Using entire test data for performance assessment	Using events in which rainfall occurs and water level rises due to flooding for performance assessment	Using events in which both rainfall occurs and water level changes for performance assessment.	Using events in which water depth overflows stream full capacity due to rainfall occurrence
Event start point	Earliest data of test data	When the water depth rises	When the water depth rises	When the water depth exceeds the stream full capacity
Event end point	Latest data of test data	When the water depth reaches peak flow	When the gradient of water discharging shows depletion curve	When the water level is lower than the stream full capacity
Classification mode	No classification	Based on both rainfall and the stream water depth data	Based on both rainfall and the stream water depth data	Based on rainfall and the stream water depth data and the stream full capacity
Focus mode	Model performance of all test data	Contribution of rainfall in water level uprising	Duration of water depth rising and then backing to the normal	True and false alarm of Overflow occurrence
Strength	Provide prediction for all weather condition	Sensitive to rainfall data	Sensitive to both rainfall and water depth data	Focusing on critical situations i.e. When overflow occurs
Weakness	The goodness of the model performance is weakened by a large part of unnecessary data	Neglecting the model performance for the falling limb in hydrograph	Lack of focus on overflow conditions	Neglecting the model performance when the water depth changes due to rainfall occurrence except for overflow

The water level-based approach (See its general steps in Algorithm 2) has two steps: it first captures water depth rising and continues until when water depth returns to the normal situation defined when water depth transforms from recession curve to the depletion curve as shown in Figure 2; Finally, an alarm-based approach which is developed based on the algorithm of the second approach (See Algorithm 3), maintains only captured water depth data in the second approach which represent the overflow condition (higher than the full capacity of catchment). It should be noted that all introduced approaches

Algorithm 1: Pseudo code of rainfall-based approach

```
Input: "Rainfall data"
For i=1 to size rainfall data
|   If the non-zero value of rainfall data is recorded continuedly
|   |   Capture data as event
|   |   Consider m time steps delay for possible rain interruption (data with zero value)
|   | Else if water level data shows falling limb in the hydrograph after rainfall occurrence
|   | Capture data as event
|   End if
End for
Output: data of recognised events
Input: "recognised events' data" and "Water depth data"
For j=1 to the total number of recognised events
|   If the water depth changes during the identified event
|   |   Confirm as a flood event and record the relevant data as an individual event
|   | Else
|   End if
End for
```

Algorithm 2: Pseudo code of water level-based approach

```
Input: Water depth data
For i=1 to size water depth data
|   If the water depth is increased continuedly
|   |   Capture as event
|   |   Consider m time steps delay for possible interruption
|   | Elseif water depth stays constant or decline
|   |   Consider as flow discharging and neglect
|   | Elseif after rising, water depth decrease
|   |   Measure ΔW_t=W_t/W_t+1 (differences of water depth)
|   |   While ΔW_t goes sharply
|   |   |   Continue to capture as the same event
|   |   End while
|   End if
End for
Output: data of recognised events
Input: "recognised events' data" and associated "Rainfall data"
For j=1 to the total number of recognised events
|   If summation rainfall data during event ~= 0
|   |   Confirm the event
|   |   Capture data of water depth for each measurement
|   End if
End for
```

Algorithm 3: Pseudo code for Alarm-based approach

```
Input: "E3: Events identified by Algorithm 2" and "Full capacity of catchment"
For i=1 to size E3
|   If water depth >= Full capacity of the catchment
|   |   Capture water depth data as an overflow data
|   |   Goes to the next event
|   End if
End for
```

3. Pilot study and model development

The suggested data classification methods are demonstrated here for a real-time AI-based flood forecasting model applied to a real-world urban catchment in the UK. The urban catchment in this pilot study is located in the London Borough of Hillingdon as shown in Figure 3. One gauging station located at river Pinn and three rainfall stations installed in RAF Northolt, Heathrow, and Iver Heath are used to collect water depth of the river and rainfall data, respectively. The developed AI-based model aims to forecast the water depth of the river in the Ruislip's urban drainage system (UDS). This pilot study was selected due to its vulnerability to frequent flooding occurrences over the Ruislip urban neighbourhoods.

The framework for development of AI-based model for real-time flood forecasting in this study is shown in Figure 4. The framework comprises three main steps including data collection, model development and model performance assessment. The first step entails data collection and preparation for two main datasets including rainfall and water depth in the river as described above. The duration of the data used for this study was continuous 11-year period from 02/08/2009 to 30/03/2020 collected from database of the Environment Agency which is accessible via public domain platforms (DEFRA, 2021). Linear regression was also used for data imputation and infilling missing data (Chen *et al.* 2021). The second step includes model development for real-time flood forecasting. The AI-based model in this study was developed based on a nonlinear autoregressive network with exogenous inputs (NARX) with two 5-node layers in the MATLAB software tool 2019a.A cross-correlation and autocorrelation were first carried out between the rainfall data and water depth data to find the best time lag between inputs (u_1=9; u_2: 10; u_3:10; v=0). Developed model used 70%, 15% and 15% of the total data for calibration, validation, and test processes, respectively. Predictions are provided for 15-minutes steps ahead, particularly the next 15-minute (t+1), 1-hour (t+4), 2-hour (t+8) and 3-hour (t+12). 10 epochs (i.e. iterations) were adjusted for training failure and 5 different iterations were repeated in each timestep.

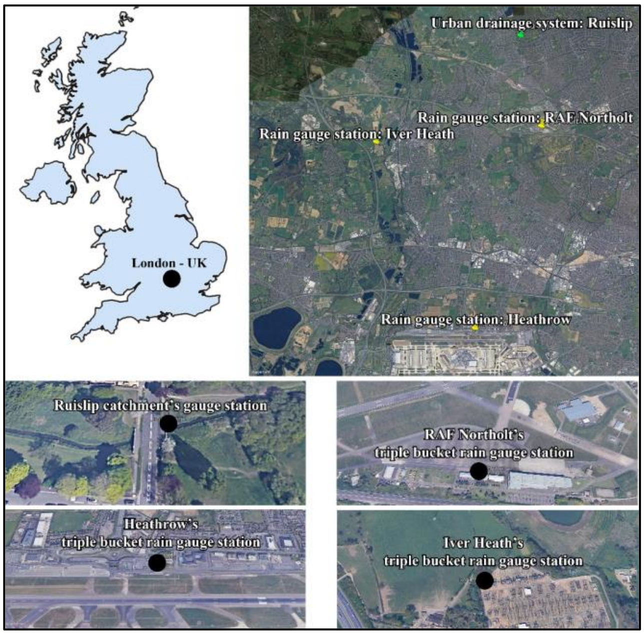


Figure 3. Location of the case study

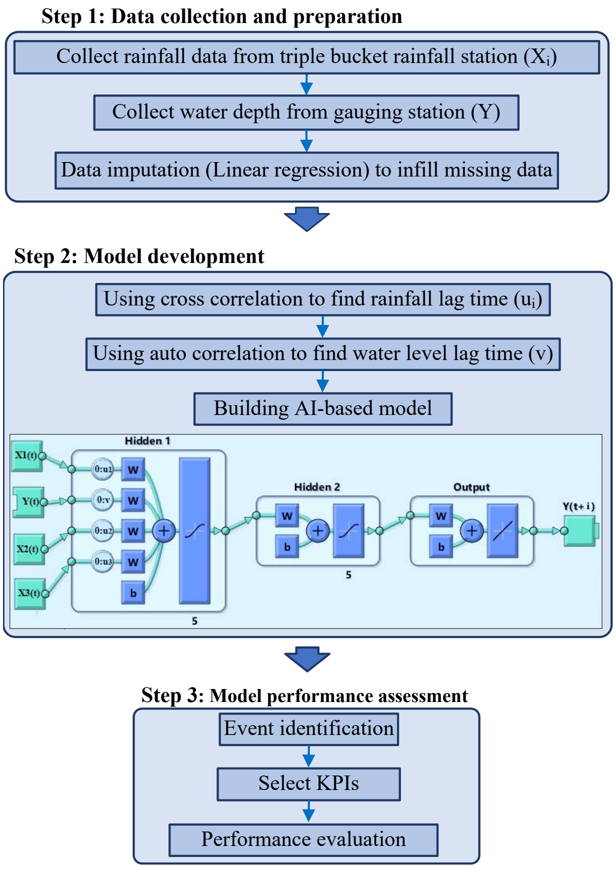


Figure 4. The methodology applied for modelling case study's urban flood forecasting

4. Results and discussion

The developed models were then used to forecast water depth at the gauging station for a few different timesteps ahead including 1 timestep (15 minutes), 4 timesteps (60 minutes), 8 timesteps (120 minutes) and 12 timesteps (180 minutes). The model performance of these predictions was evaluated by using two known KPIs i.e., RMSE (Root Mean Square Error), and NSE (Nash Sutcliffe Efficiency), which are shown in Table 3. It can be seen from Table 3 that among all test data used for performance evaluation of model, only 1.33 to 5.72% of the total data (112,109) are identified as flood events within the three data classification methods. In other words, 94% of the total data that are used in the conventional method for the model performance assessment are related to non-flood data (i.e., one of the three states in Figure 1) and hence are redundant for the model performance assessment. As can be seen, figures for both RMSE and NNSE in the conventional method are close to non-flood event data for other data classification methods as thus data account for majority of the database. The good performance of indicators for non-flood events confirms that high accuracy for predictions of water depth within those periods that are basically unnecessary for the modelling purposes.

The performance of the rainfall-based and water level-based methods seems similar while the alarm-based method is the worst with respect to all indicators (i.e. RMSE is double and NNSE is lower by 3% in the overflow approach in comparison to other introduced approaches). This means that model could not outperform overflow conditions as well as water depth variation when flood events do not cause overflow.

Table 3. KPIs of AI-based model for one step ahead of prediction

Data classification Method	Number (%) of test data used for performance assessment	Average of performance indicators for all trials			
		RMSE (mm)		NSE (%)	
		FE	UE	FE	UE
Conventional	112,109 (100%)	4.07		99.82	
Rainfall-based	3,287 (2.93%*)	10.92	3.86	99.40	99.83
Water level-based	6,416 (5.72%*)	10.68	3.67	99.24	99.86
Alarm-based	1,491 (1.33%*)	18.27	3.88	97.20	99.86

Performance assessment of AI-based model was also carried out for different lengths of prediction, as shown in Figure 5. Generally, RMSE and NSE became worst for longest lead time. However, differences were increased for RMSE rather than NSE. RMSE changed from 4.07 to 33.45 in conventional approach, whereas they raised from about 11 to near 37 for both rainfall based and water level-based approaches. These defences show accuracy of model was significantly affected by unnecessary data in 1-step ahead but then it blurred in 12-step ahead mainly because of inaccuracy of flood prediction in longest time steps. NSE performance, however, shows that model tracking decreased only 10% for all approaches form 1-step to 12-step ahead.

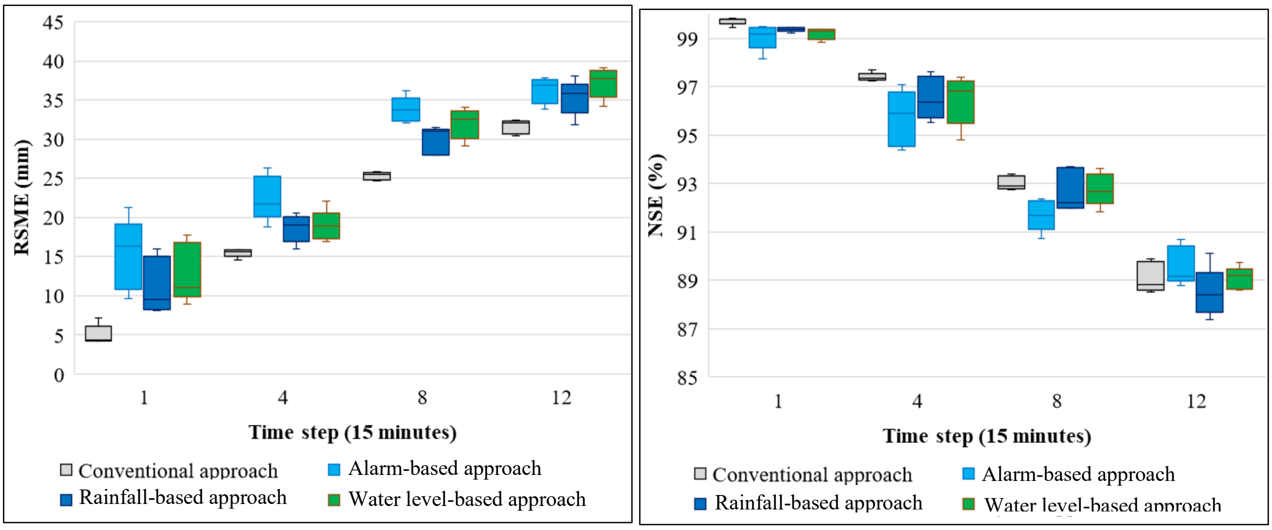


Figure 5. Model performance based on each of the identified approaches for four different lead-time prediction

5. Conclusions

The present study devotes to dealing with the high demands on providing realistic performance evaluation of time-series AI-based models through proposing data classification approaches for flood event identification. Data classification approaches show that a huge part of the test database may not represent flood events (95% of the entire test database of case study, for instance), which demonstrate the importance of the screen method for removing this unnecessary data from performance evaluation. Additionally, the proposed data classification approaches revealed that flood event measurement can provide a more realistic model performance evaluation. While the conventional approach may reflect a suitable model's ability in providing accurate estimations, new approaches showed that many of performed ability is not associated with the wet weather condition and flooding occurrence. In total, three classification approaches are proposed to cover all aspects of flood monitoring and early warning systems. While the rainfall-based approach focuses on the contribution of rainfall in water depth uprising, the water level-based approach emphasises on hydraulic characteristics of the chamber. The Alarm-based approach also detects the ability of the model on accurate estimation of overflow conditions. Each of these approaches can be used for performance evaluation based on the goals of modelling. For the case study, there is clear differences between translating ability of model in urban flood forecasting, when entire data are considered in comparison to when other approaches are introduced. Differences are highlighted in wider time steps, especially in RMSE. While all performances are acceptable, differences shows that event identification can provide more realistic examination about ability of modelling. Finally, while demonstrated case study approved the potential power of the proposed framework, it is worth to conduct further inclusive research to open a new door for adopting it in the time-series AI-based flood forecasting models performance and clarifying both benefits and drawbacks.

References

Accarino, G., Chiarelli, M., Fiore, S., Federico, I., Causio, S., Coppini, G., Aloisio, G. (2021). A multi-model architecture based on Long Short-Term Memory neural networks for multi-step sea level forecasting. Future Generation Computer Systems, 124, pp. 1-9.

Adikari, K., Shrestha, S., Ratnayake, D., Budhathoki, A., Mohanasundaram, S. Dailey, M. (2021). Evaluation of artificial intelligence models for flood and drought forecasting in arid and tropical regions. Environmental Modelling & Software, 144: 105136.

Chen, C., Hui, Q., Xie, W., Wan, S., Zhou, Y., Pei, Q. (2021). Convolutional Neural Networks for forecasting flood process in Internet-of-Things enabled smart city. Computer Networks, 186, 107744.

Daal, P., Gruber, G., Langeveld, J., Muschalla, D., Clemens, F. (2017). Performance evaluation of real time control in urban wastewater systems in practice: Review and perspective. *Environmental Modelling & Software*, 95, pp. 90-101.

Darabi, H., Haghighi, A., Mohamadi, M. Rashidpour, M. Ziegler, A., Hekmatzadeh, A., Kløve, B. (2020).

Urban flood risk mapping using data-driven geospatial techniques for a flood-prone case area in Iran. *Hydrology Research* 51 (1): 127–142.

Department for Environment, Food and Rural Affairs (DEFRA). (2021). *DEFRA Official Website* [Online] Available at http://Environment.data.gov.uk, [Accessed 12 Jan. 2022].

Huang, X., Li, Y., Tian, Z., Ye, Q., Ke, Q., Fan, D., Mao, G., Chen, A., Liu, J. (2021). Evaluation of short-term streamflow prediction methods in Urban river basins. Physics and Chemistry of the Earth, 123: 103027.

Kourtis, I., Tsihrintzis, V. (2021). Adaptation of urban drainage networks to climate change: A review. *Science of The Total Environment*, 771, Doi.org/10.1016/j.scitotenv.2021.145431.

Lin, Y., Wang, D., Wang, G., Qiu, J., Long, K., Du, Y., Xie, H., Wei, Z., Shangguan, W., Dai, Y., (2021). A hybrid deep learning algorithm and its application to streamflow prediction. Journal of Hydrology, 601: 126636.

Mosavi, A., Ozturk P., Chau K. (2018). Flood Prediction Using Machine Learning Models: Literature Review. *Water*, 10(11), 1536.

Piadeh, F., Behzadian, K., Alani, A.M. 2022. A critical review of real-time modelling of flood forecasting in urban drainage systems. *Journal of Hydrology*, 607: 127476, doi.org/10.1016/j.jhydrol.2022.127476.

Thrysøe, C., Arnbjerg-Nielsen, K., Borup, M. (2019). Identifying fit-for-purpose lumped surrogate models for large urban drainage systems using GLUE. *Journal of Hydrology*, 568, pp. 517-533.

Zounemat-Kermani, M., Matta, E., Cominola, A., Xia, X., Zhang, Q., Liang, Q., Hinkelmann, R. (2020). Neurocomputing in surface water hydrology and hydraulics: A review of two decades retrospective, current status and future prospects. *Journal of Hydrology*, 588, 125085.