https://doi.org/10.71573/m854dw13

Water level prediction in urban drainage systems using explainable deep learning models

Seith N. Mugume[1,*] https://orcid.org/0000-0002-8289-0099, *Dorothy Pamela Adeke*[1]

[1]*Makerere University, Department of Civil and Environmental Engineering, P.O. Box 7062, Kampala*
**Corresponding author email: smugume@gmail.com; seith.mugume@mak.ac.ug*

Abstract

Accurate water level prediction in existing urban drainage systems (UDSs) is critical for reliable forecasting of pluvial flooding impacts and reduction of flood-related damages in cities. Conventional physically based urban drainage modelling approaches are constrained by the need for extensive hydro-meteorological, drainage network and surface terrain data and high computational demands. In this research, more computationally efficient Machine Learning based Feedforward Neural Network (FFNN), multi-head Convolutional Neural Network (CNN) and 1D-CNN models were developed and applied to simulate water levels at a bridge crossing downstream of an existing UDS in Kampala City. The study results suggested that the multi-head CNN Deep Learning model resulted in more superior predictive performance (NSE, RMSE and MAE of 0.564, 0.208, and 0.091) when compared to the physically based PCSWMM model (NSE, RMSE, and MAE of 0.505, 0.221 and 0.098). Furthermore, the SHapley Additive exPlanations (SHAP) approach was applied to explain the underlying processes in the developed ML models and to determine the most influential model parameters. The research demonstrates that explainable Deep Learning models can reliably simulate water levels in UDSs, and provide a robust basis for development of real-time pluvial flood early warning systems in data-scarce cities.

Highlights

- Comparative evaluation of Deep Learning models in simulating water levels in UDSs.
- SHAP approach used to analyse feature importance and to enhance model interpretability.
- Multi-head CNN model resulted in more superior water level predictive performance.

Introduction

Globally, urban flooding is a persistent challenge that is being exacerbated by climate change, rapid urbanisation, component failures, sediment and solid waste deposition and deterioration of existing urban drainage infrastructure (Adeke & Mugume, 2025; MacAfee & Löhr, 2024; Olsson et al., 2025). Accurate water level simulation and prediction is critical for development of pluvial flood early warning systems that can contribute to minimisation of resulting flood related damages (Chen et al., 2023; Hieu et al., 2023; Liu et al., 2024). Conventional approaches for prediction of urban flooding impacts include physically based one-dimensional (1D), 1D sewer and 1D surface flow (1D-1D), two-dimensional (2D) surface flow and coupled 1D-2D modelling (Ghimire et al., 2013; Liu et al., 2024; Mugume, Kibibi, et al., 2024; Webber et al., 2019). The use of physically-based coupled 1D-2D or 2D surface flow models in undertaking continuous or real-time simulations is constrained by the need for accurate representation of urban drainage network and 2D surface terrain data, high computational resource requirements, inherent model uncertainties and insufficient observed data for model calibration/validation (Ghimire et al., 2013; Moy de Vitry & Leitão, 2020). To address the shortcomings of conventional physically based models, Artificial Intelligence (AI) and Machine Learning (ML) based models such as Artificial Neural Networks (ANNs) and hybrid ANNs have been developed and applied to simulate flooding impacts in cities.

However, the performance of ANN models is constrained by unavailability of relatively long input datasets with sufficient number of peaks for effective model calibration, leading to their inability to establish strong patterns (Mugume, Murungi, et al., 2024). In more recent studies, Deep Learning (DL) models such Graph Neural Networks (GNN), Gated Recurrent Unit (GRU), Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNNs) and their hybrids models have been applied to achieve reliable and fast prediction of water levels and flooding impacts (Garzón et al., 2024; Hieu et al., 2023; Rasha Jamal Hindi, 2025). In this research, sub-daily rainfall and water level measurements were undertaken from March 2024 to March 2025 in the upper Lubigi urban drainage system (UDS) in Kampala City, Uganda. The observed rainfall and water level data sets were utilised as input and target data respectively for development and comparative evaluation of the performance of a coupled 1D-2D PCSWMM, FFNN and two Deep Learning models utilising a 1D-CNN and a multi-head CNN in prediction of water levels at Hoima Road bridge crossing, located at the outfall of the UDS. The study addresses a critical research gap of developing novel explainable ML based approaches for fast simulation of flooding impacts and presents a promising basis for development of pluvial flood early warning systems in data scarce cities.

Methodology

Data sets and data pre-processing

The datasets used in this research include: high temporal resolution rainfall data which was collected using a HOBO Data logging automatic tipping bucket rain gauge RG3-M) installed within the Lubigi faecal sludge and waste water treatment in Kampala (Olsson et al., 2025). Other research data sets included observed sub-daily water level measurements (WLMs) undertaken a Hoima Road bridge crossing, 30 m x 30 m Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (DEM), Sentinel-2 satellite imagery and existing urban drainage network data (open channels, cross culverts, storage units, pipes, inlets) that was provided by Kampala Capital City Authority. The rainfall data collected using a HOBO Data logging automatic tipping bucket rain gauge RG3-M. These datasets were used to undertake catchment delineation and assessment of imperviousness levels at a sub-catchment level. model utilised a digital elevation model (DEM), Sentinel-2 satellite imagery and drawings from Kampala Capital City Authority (KCCA) for catchment delineation, percentage imperviousness calculations and drainage infrastructure details integration respectively. The water level measurements at the downstream location of the UDS were taken twice a day and immediately after occurrence of a rain event. With the high-resolution rainfall data, the accumulated rainfall and timing features were used to align the rainfall input data to the water level measurements. Furthermore, the inputs for the ML models were derived from the high-resolution rainfall data and included the accumulated rainfall (RA) between water level measurements, the rainfall duration (RD) and the elapsed time since the end of each rainfall event (TR). These inputs were utilised in the ML models to learn from the rainfall characteristics before each water level observation.

Physically based and Machine Learning (ML) modelling

As a first step, a 1D hydraulic model was developed in PCSWMM model based on an approach detailed in Mugume *et al.*, (2024). The developed model was calibrated using 16 rainfall events from 3rd to 23rd April 2024 for the wet season and 6 rainfall events from 1st to 31st August 2024 for the dry season. The model was thereafter validated with 45 water level measurements (WLM) from 6th to 26th March 2024 for the wet season and 57 WLM from 1st to 31st January 2025 for the dry season. Three Machine Learning (ML) models namely; the FFNN and two Deep Learning (Multi-head CNN, and 1D-CNN) models were developed in Python Programming Language, trained, validated and tested. The ML models were trained using March to August 2024 rainfall and water level data. This period covers both the first wet season and the dry season after, that is March, April, and May (MAM) for the rainy season and June, July and August (JJA) for the dry season. The models were then used to simulate downstream water levels for the secondary wet season (September, October, and November i.e. SON) and the following dry season (December, January, and February i.e. DJF). Each of the ML models was chosen due to its unique capabilities. The FFNN model architecture consists of an input layer, one or more hidden layers, and an output layer. It was chosen as a baseline due to its simplicity and ability to accurately simulate complex non-linear processes without the need for detailed knowledge on underlying hydrological processes (Khalil et al., 2022; Mugume, Murungi, et al., 2024). The 1D-CNN model was selected due to its capability to capture dependencies through convolutional filters that move along the time dimension, making it well-suited for time series data (Cacciari & Ranfagni, 2024; Tang et al., 2020). In addition, the multi-head CNN model uses multiple convolutional filter sizes at the same time, that enable it to capture features across different rainfall durations and scales (Yan et al., 2020). Furthermore, the rectified linear (ReLu) activation function was utilised in the developed ML models. The description of the underlying model structures for the FFNN, multi-head CNN and 1D-CNN models is described in literature (Cacciari & Ranfagni, 2024; Jun et al., 2024; Mugume, Murungi, et al., 2024; Xu et al., 2024).

Results and discussion

The study results suggest that the developed multi-head CNN (NSE, RMSE and MAE of 0.564, 0.208 and 0.091) is more superior in prediction of downstream water levels when compared to the FFNN (NSE, RMSE and MAE of 0.556, 0.210 and 0.098) and the physically based PCSWMM model (NSE, RMSE and MAE of 0.505, 0.221 and 0.098). In addition, the performance of the multi-head CNN for the entire period of monitoring was comparable to results obtained using the 1D-CNN (NSE, RMSE and MAE of 0.586, 0.203 and 0.091) and physically-based PCSWMM model (NSE, RMSE and MAE of 0.505, 0.221 and 0.098) (Table 1, Figure 1 and Figure 2). The study results also suggest superior performance of the multi-head CNN and 1D-CNN in simulating low flows during the December, January and February (DJF) season (Table 1). The scatter plots in Figure 3 further illustrate that globally, the multi-head CNN resulted in more superior water level predictive performance. This is attributed to the multi-layer structure that enables multiple convolutional layers to work in parallel in learning diverse features and patterns in the input data over different lengths and temporal resolutions (Khan & Ahmad, 2021) and the diverse kernel sizes of the convolutional filters of each branch (Xu et al., 2024).

Table 1: Performance metrics of the developed PCSWMM, FFNN, multi-head CNN and 1D-CNN models

	Training and validation			Prediction (SON)			Prediction (DJF)		
Model	NSE	RMSE	MAE	NSE	RMSE	MAE	NSE	RMSE	MAE
PCSWMM	0.505	0.221	0.098	0.467	0.264	0.131	0.560	0.150	0.054
FFNN	0.641	0.138	0.064	0.436	0.276	0.143	0.677	0.128	0.057
Multi-head CNN	0.641	0.136	0.07	0.445	0.275	0.139	0.693	0.125	0.049
1D-CNN	0.742	0.117	0.056	0.532	0.252	0.118	0.726	0.118	0.053

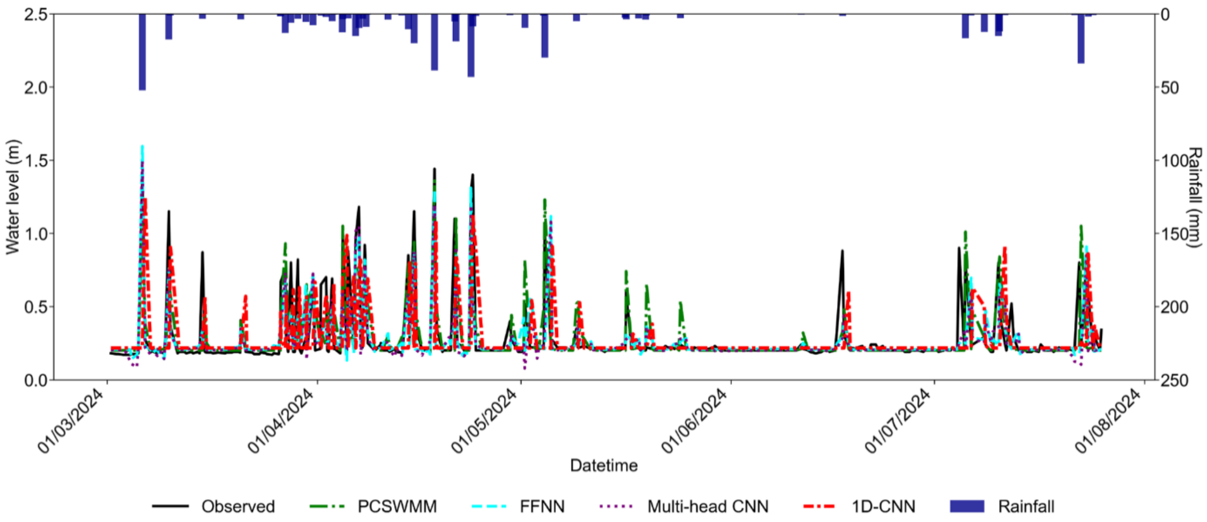

Figure 1: A plot of observed and simulated water levels during model training and validation

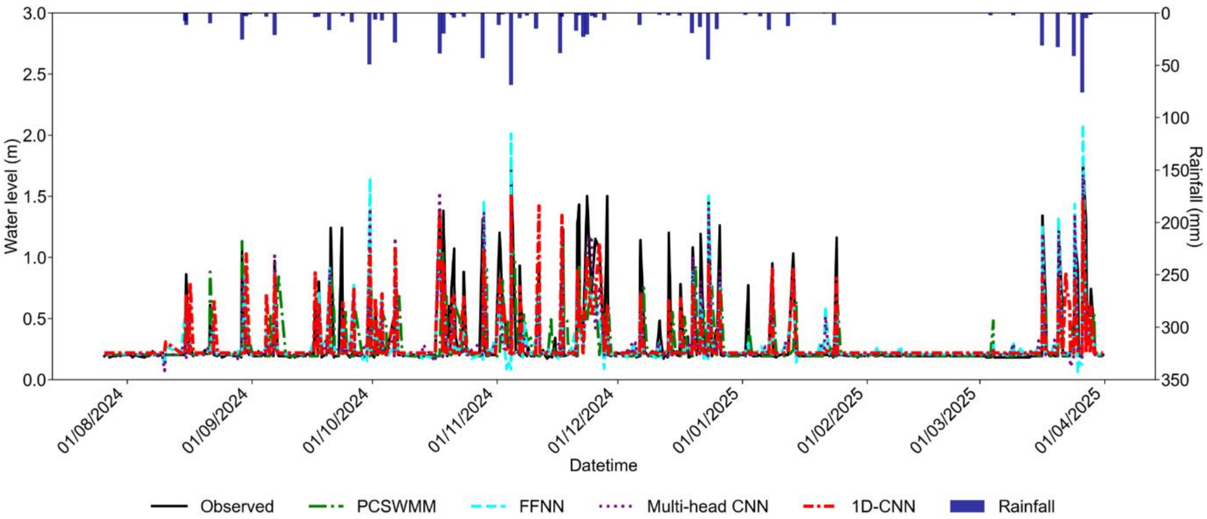

Figure 2: A plot of observed and simulated water levels during testing and simulation

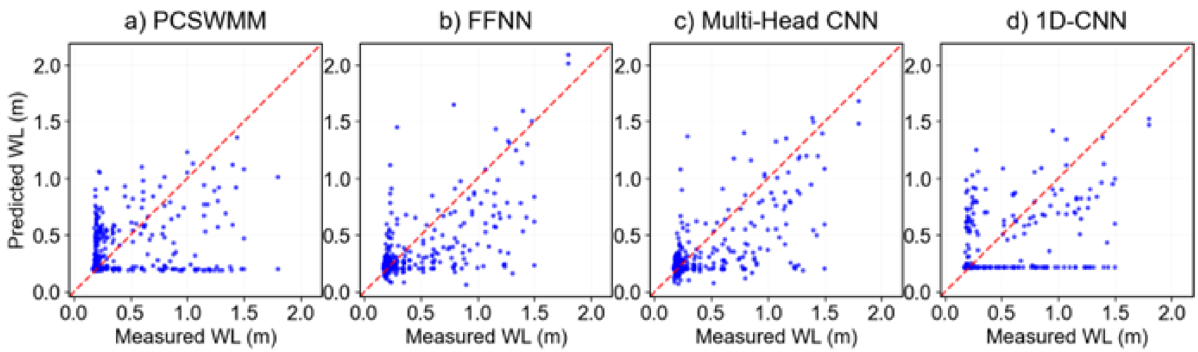

Figure 3: Scatter plots of observed and simulated water levels obtained using (a) PCSWMM, (b) FFNN, (c) Multi-head CNN and (d) 1D-CNN models

In addition, the SHapley Additive exPlanations (SHAP) game theoretic approach was applied to explain the underlying processes in the considered deep ML models and to determine the most influential model parameters (Figures 4, 5 and 6). In this study, RA_i is used to denote the accumulated rainfall at time, t when a water level measurement was taken, RD_i the rainfall duration, TR_i the elapsed time since the end of the rainfall event and i the time step number. For all the models, more than one input was used during simulation of the downstream water levels. At a given instance, 4- or 5-time steps were used in the simulation. The rainfall amount, duration and the elapsed time since the end of the rainfall event are parameters that are known to affect the water levels in UDSs (Joo *et al.*, 2014; Laouacheria *et al.*, 2019; Ma *et al.*, 2024).

Figures 4 a), b) and c) present the SHAP analysis results for the water level instance prediction for the 26th March 2025 extreme rainfall event, with a total rainfall depth of 76 mm that caused exceedance flows and extensive flooding in Kampala (Kiiza, 2025). The plot shows the contributions of each input parameter to the final predicted water level starting from the base value (average value). The plots illustrate the importance and magnitude of influence of the RA_i, RD_i and TR_i values at different time steps. The magnitude is illustrated by the SHAP values indicated in or beside the red and blue bars. The red bars indicate the features that increase the predicted from the base value while the blue bars show those that decrease. The final simulation is a sum of the base value and the contributed SHAP values from the input parameters. A high value of RA1 (76 mm) for a given RD1 (16,200 s) in Figure 4 a) and b) increases the prediction while a high TR1 (13,860 s) decreases the prediction in Figure 4 d). The low value of RA4 (0.4 mm) and RA2 (1 mm) contributed minimally as compared to RA1 in Figure 4 c). Figure 4 d) illustrates the prediction for a low intensity rainfall event day with RA1 (1 mm), RD1 (4,500 s), and TR3 (13,860 s) leading to the final prediction equal to the base value.

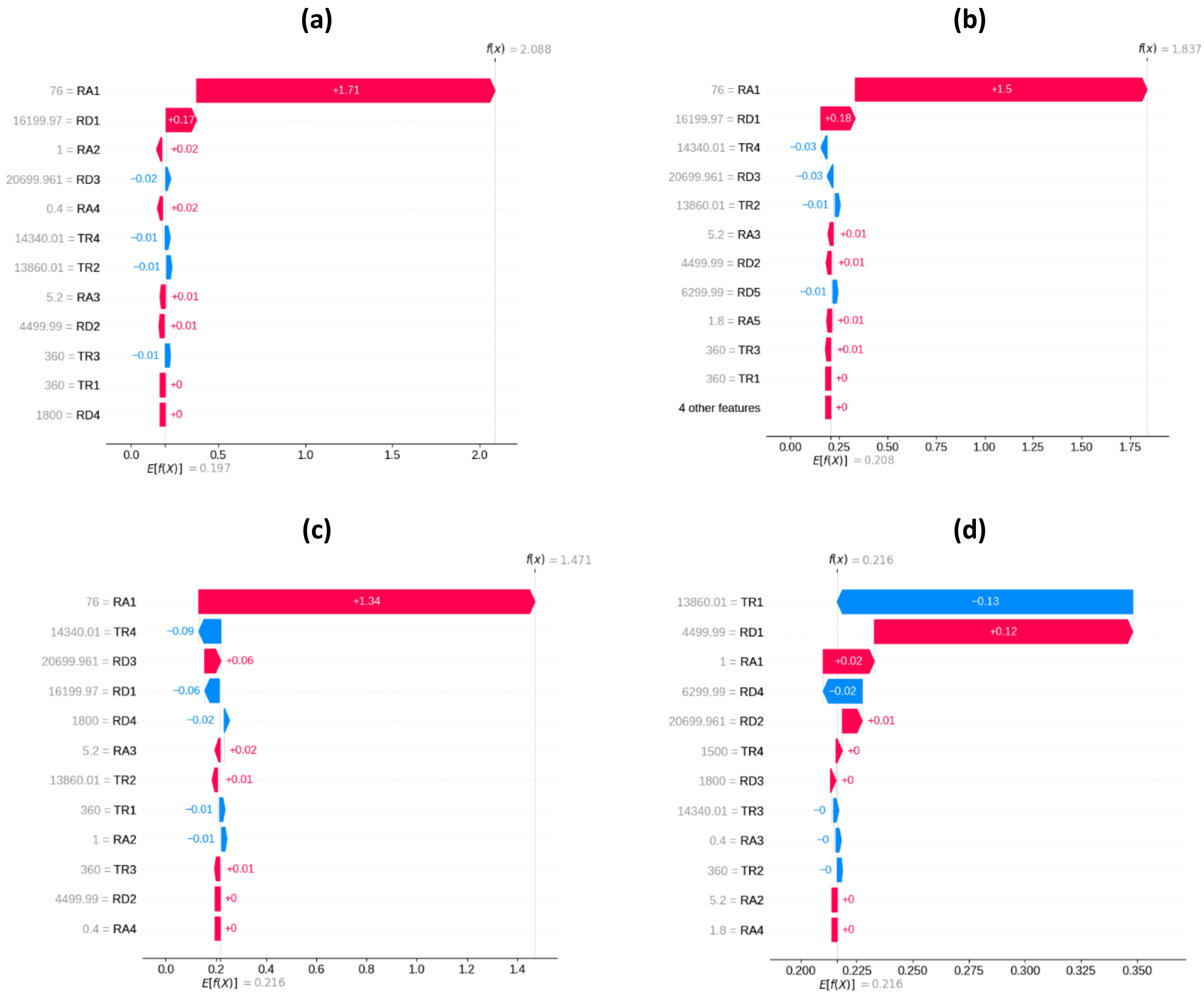

Figure 4: Instance prediction in (a) FFNN (b) Multi-head CNN (c) 1D-CNN and (d) second instance prediction in the 1D-CNN

The temporal categories considered included: “No-Rain” for total rainfall is less than 0.2 mm, “Early-Peak” in which the peak rainfall occurs in the first third of the sequence and “Late-Peak”, in which the peak rainfall occurs in the final third of the sequence (Figure 5). The rainfall patterns were categorised as light, intense short duration and long drizzle rainfall events. Light rainfall events were those in which the total rainfall was less than the 25th percentile. In addition, intense short duration events were those in which the total rainfall is more than the 75th percentile with an average duration less than the median value, while the long-drizzle events were those with the total rainfall less than the 75th percentage with an average duration greater than twice the median value. The moderate rainfall, which represented any event that did not fall in the aforementioned categories, was not represented because it gave similar plots to the short duration intense rainfall events with a RD1 feature having a slightly higher mean SHAP value.

The results of the SHAP analysis also suggested that the water level at time, *t* is influenced by rainfall amounts from the previous 4 to 5 rainfall events and their individual duration. Furthermore, the SHAP results suggest that effect of rainfall amount and duration diminish over time and that the most recent rainfall amount and duration was the most influential factor for the resulting downstream water levels. In addition, the elapsed time since the end of a rainfall event also significantly influenced downstream water levels as the time from the end of a given rainfall event increased. The results are agreement with the findings of Xiang et al., (2024) and Jiang et al., (2022) and who identified recent rainfall and antecedent rainfall (which is linked to excessive soil moisture) as the main drivers of peak flooding events. Lastly, the study findings suggest that the developed Deep Learning models can effectively predict the effect of different rainfall characteristics on the downstream water levels at varying periods.

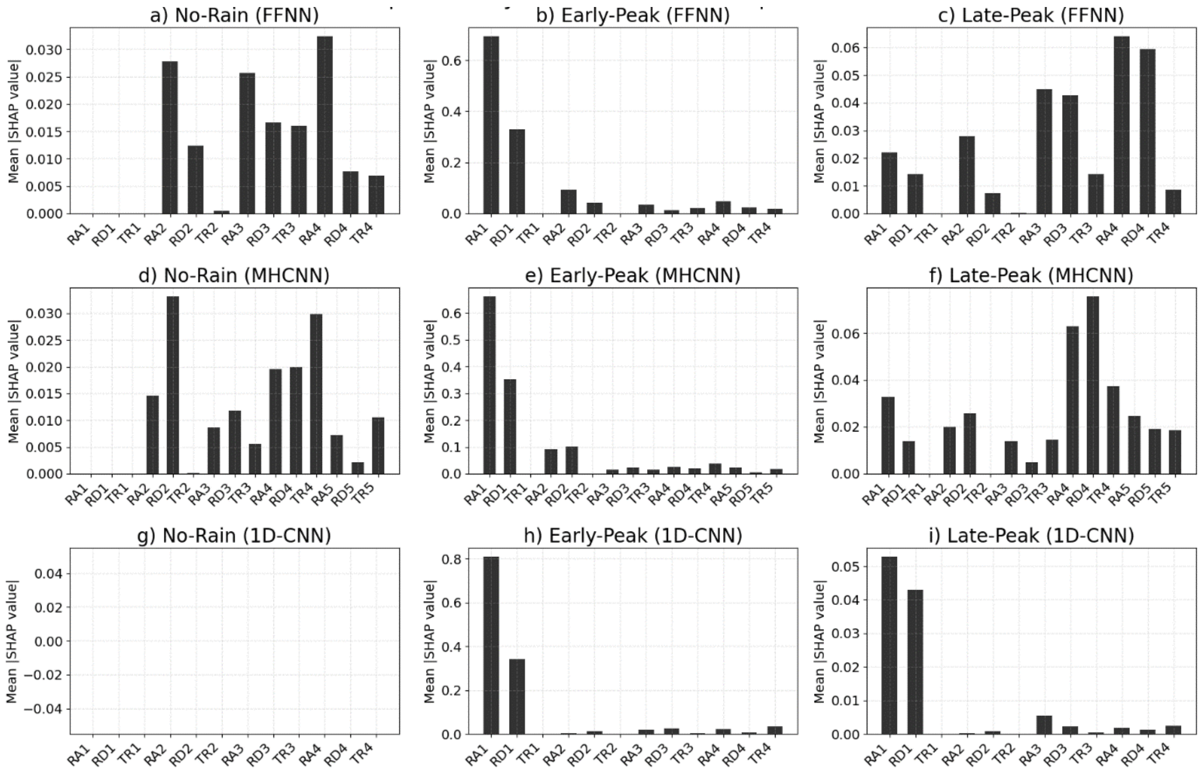

Figure 5: Analysis of the developed ML model capability in capturing temporal rainfall patterns: (a) – (c) FFNN; (d) – (f) Multi-Head CNN and (g) – (i) 1D CNN.

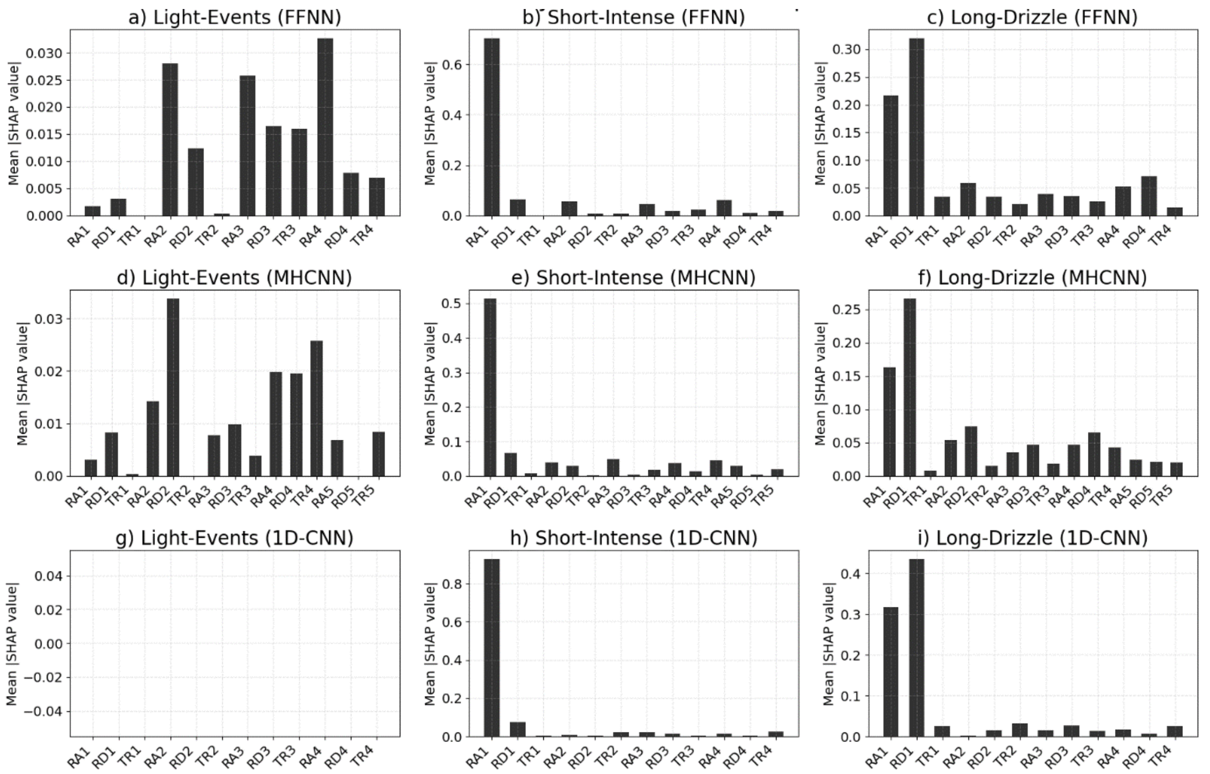

Figure 6: Analysis of the developed ML model capability in capturing rainfall event patterns: (a) – (c) FFNN; (d) – (f) Multi-Head CNN and (g) – (i) 1D CNN.

Conclusions and future work

In this research, three explainable Machine Learning (ML) models that included the Feedforward Neural Network (FFNN) and two Deep Learning (DL) models utilising the multi-head Convolutional Neural Network (CNN) and a 1D-CNN were developed and applied to simulate water levels at a bridge crossing downstream of an existing UDS in Kampala City. The study findings suggest that explainable DL models such as the multi-head Convolutional Neural Networks can reliably predict water levels in existing UDSs in data scarce cities with considerable accuracy due to their reliability in capturing the effect of varying rainfall characteristics on the resulting downstream water levels. For all considered ML models, accumulated rainfall showed the strongest influence on increases in the downstream water levels, while the elapsed time since the end of a given rainfall event showed the strongest influence on decreases in the predicted downstream water levels. Future research that utilises explainable hybrid ML models and transformers (which enhance the representation of spatial and temporal dependences) for prediction of pluvial flooding thresholds using future rainfall derived from downscaled global climate models (GCMs) is recommended. The presented study demonstrates that explainable Deep Learning models can reliably simulate water levels in UDSs, and provide a robust modelling framework for future development of real-time pluvial flood early warning systems in data-scarce cities.

References

Adeke, D. P., & Mugume, S. N. (2025). A methodology for development of flood-depth-velocity damage functions for improved estimation of pluvial flood risk in cities. *Journal of Hydrology*, *653*(132736), 1–17. https://doi.org/https://doi.org/10.1016/j.jhydrol.2025.132736

Cacciari, I., & Ranfagni, A. (2024). Hands-On Fundamentals of 1D Convolutional Neural Networks—A Tutorial for Beginner Users. *Applied Sciences (Switzerland)*, *14*(18), 1–26. https://doi.org/10.3390/app14188500

Chen, J., Huang, Y., Wu, T., & Yan, J. (2023). A WaveNet-based convolutional neural network for river water level prediction. *Journal of Hydroinformatics*, *25*(6), 2606–2624. https://doi.org/10.2166/hydro.2023.174

Garzón, A., Kapelan, Z., Langeveld, J., & Taormina, R. (2024). Transferable and data efficient metamodeling of storm water system nodal depths using auto-regressive graph neural networks. *Water Research*, *266*(122396), 1–14.

https://doi.org/10.1016/j.watres.2024.122396

Ghimire, B., Chen, A. S., Guidolin, M., Keedwell, E. C., Djordjević, S., & Savić, D. A. (2013). Formulation of a fast 2D urban pluvial flood model using a cellular automata approach. *Journal of Hydroinformatics*, *15*(3), 676–686. https://doi.org/10.2166/hydro.2012.245

Hieu, T. T., Chieu, T. Q., Quang, D. N., & Hieu, N. D. (2023). Water level prediction using deep learning models. *River*, *2*(4), 468–479. https://doi.org/10.1002/rvr2.63

Jiang, S., Bevacqua, E., & Zscheischler, J. (2022). River flooding mechanisms and their changes in Europe revealed by explainable machine learning. *Hydrology and Earth System Sciences*, *26*, 6339–6359. https://doi.org/10.5194/hess-26-6339-2022

Joo, J., Lee, J., Kim, J. H., Jun, H., & Jo, D. (2014). Inter-event time definition setting procedure for urban drainage systems. *Water*, *6*(1), 45–58. https://doi.org/10.3390/w6010045

Jun, S., Jung, D., & Lansey, K. (2024). Three-dimensional convolutional neural network for leak detection and localization in smart water distribution systems. *Water Research X*, *25*(August), 100264. https://doi.org/10.1016/j.wroa.2024.100264

Khalil, K., Kumar, A., & Bayoumi, M. (2022). Reconfigurable Hardware Design Approach for Economic Neural Network. *IEEE Transactions on Circuits and Systems II: Express Briefs*, *PP*, 1. https://doi.org/10.1109/TCSII.2022.3191342

Khan, Z. N., & Ahmad, J. (2021). Attention induced multi-head convolutional neural network for human activity recognition. *Applied Soft Computing*, *110*(107671), 1–9. https://doi.org/10.1016/j.asoc.2021.107671

Kiiza, C. (2025, March 26). Kampala Floods: Death Toll Rises to 7 as US Embassy Staff Drowns. *ChimpReports*. https://chimpreports.com/kampala-floods-death-toll-rises-to-7-as-u-s-embassy-staff-drowns/ (Accessed 30th July 2025)

Laouacheria, F., Kechida, S., & Chabi, M. (2019). Modelling the impact of design rainfall on the urban drainage system by Storm Water Management Model. *Journal of Water and Land Development*, *40*(1), 119–125. https://doi.org/10.2478/jwld-2019-0013

Liu, B., Li, Y., Feng, X., & Lian, P. (2024). BEW-YOLOv8: A deep learning model for multi-scene and multi-scale flood depth estimation. *Journal of Hydrology*, *645*(132139), 1–12. https://doi.org/10.1016/j.jhydrol.2024.132139

Ma, T., Kim, J. S., Jun, C., Moon, Y. Il, & Moon, H. (2024). Optimizing urban flood management: enhancing urban drainage system efficiency under extreme rainfall events. *Journal of Hydroinformatics*, *26*(11), 2704–2719. https://doi.org/10.2166/hydro.2024.067

MacAfee, E. A., & Löhr, A. J. (2024). Multi-scalar interactions between mismanaged plastic waste and urban flooding in an era of climate change and rapid urbanization. *Wiley Interdisciplinary Reviews: Water*, *11*(e1708), 1–13. https://doi.org/10.1002/wat2.1708

Moy de Vitry, M., & Leitão, J. P. (2020). The potential of proxy water level measurements for calibrating urban pluvial flood models. *Water Research*, *175*, 1–14. https://doi.org/10.1016/j.watres.2020.115669

Mugume, S. N., Kibibi, H., Sorensen, J., & Butler, D. (2024). Can Blue-Green Infrastructure enhance resilience in urban drainage systems during failure conditions? *Water Science and Technology*, *89*(4), 915–944. https://doi.org/10.2166/wst.2024.032

Mugume, S. N., Murungi, J., Nyenje, P. M., Sempewo, J. I., Okedi, J., & Sörensen, J. (2024). Development and application of a hybrid artificial neural network model for simulating future stream flows in catchments with limited in situ observed data. *Journal of Hydroinformatics*, *26*(8), 1944–1969. https://doi.org/10.2166/hydro.2024.066

Olsson, H., Mugume, S. N., & Sorensen, J. (2025). Modelling the effect of sediment and solid waste deposition on pluvial flood risk using in situ high-resolution rainfall and water level measurements. *Water Science and Technology*, *91*(6), 757–782. https://doi.org/https://doi.org/10.2166/wst.2025.039

Rasha Jamal Hindi. (2025). Enhancing River Water Level Prediction Using CNN with Attention and Squeeze-and-Excitation Blocks. *International Journal of Scientific Research in Science, Engineering and Technology*, *12*(3), 1248–1269. https://doi.org/10.32628/ijsrset2512174

Tang, W., Long, G., Liu, L., Zhou, T., Jiang, J., & Blumenstein, M. (2020). *Rethinking 1D-CNN for Time Series Classification: A Stronger Baseline*. https://doi.org/10.48550/arXiv.2002.10061

Webber, J. L., Fu, G., & Butler, D. (2019). Comparing cost-effectiveness of surface water flood management interventions in a UK catchment. *Journal of Flood Risk Management*, *12*(2), 1–12.

Xiang, X., Guo, S., Cui, Z., Wang, L., & Xu, C. Y. (2024). Improving flood forecast accuracy based on explainable convolutional neural network by Grad-CAM method. *Journal of Hydrology*, *642*(131867), 1–14. https://doi.org/10.1016/j.jhydrol.2024.131867

Xu, Y., Pan, Q., Wang, Z., & Hu, B. (2024). A Novel Trajectory Prediction Method Based on CNN, BiLSTM, and Multi-Head Attention Mechanism. *Aerospace*, *11*(10), 1–19. https://doi.org/10.3390/aerospace11100822

Yan, Q., Ji, F., Miao, K., Wu, Q., Xia, Y., & Li, T. (2020). Convolutional Residual-Attention: A Deep Learning Approach for Precipitation Nowcasting. *Advances in Meteorology*, *2020*. https://doi.org/10.1155/2020/6484812