CCC 2025 **Proceedings of the Creative Construction Conference** (2025)
Edited by: Miroslaw J. Skibniewski & Miklós Hajdu & Žiga Turk
https://doi.org/10.22260/CCC2025/0050

COMPARATIVE ANALYSIS OF SHORT-TERM WATER LEVEL FORECASTING IN UNGAUGED RIVER SYSTEMS

Younga Shin, Kichang Choi, Yeonjoo Kim, **Hongjo Kim**
Yonsei University, Seoul, Republic of Korea

Abstract

The intensification of extreme rainfall events due to climate change has significantly increased flood risks, with levee overtopping emerging as a major cause of catastrophic failures. Accurate short-term water level forecasting is therefore essential, particularly in ungauged river systems where monitoring infrastructure is limited. However, existing deep learning models heavily rely on long-term observations from gauged sites, while physics-based models require dense, high-quality input data—both of which constrain real-time applicability in data-scarce regions. To address these challenges, we propose DeepCreek, a lightweight MLP-based forecasting model adapted from TSMixer, designed for short-term water level forecasting in ungauged basins. Utilizing nationwide rainfall scenario-based datasets comprising eight key hydrometeorological variables across 232 observation stations in South Korea, and employing a spatially segregated evaluation strategy, DeepCreek achieved an RMSE of 1.0987m and a MAPE of 15.43% for six-hour-ahead predictions under extreme rainfall scenarios. The model consistently outperformed conventional approaches on both the rainfall scenario and extreme rainfall scenario test sets, demonstrating robust performance even under noisy real-world conditions without explicit outlier removal. These results underscore DeepCreek's potential as a practical and scalable AI solution for real-time flood forecasting in regions with limited hydrological monitoring infrastructure.

Keywords: deep learning, rainfall scenario, time-series data, ungauged basins, water level forecasting.

Peer-review under responsibility of the scientific committee of the Creative Construction Conference 2025.

1. Introduction

1.1. Research background

Climate change has increased the frequency and intensity of extreme rainfall, raising concerns over the stability of flood protection systems such as levees. Among various failure mechanisms, overtopping has been consistently identified as one of the most critical and frequent causes [1]. It accounted for over 68% of levee breaches across nine countries [2]. In South Korea, the proportion of overtopping-related failures increased from 39.6% of 758 cases between 1987 and 2003 to 69% during the 2006 floods [3-4], indicating an escalating trend. This underscores the urgent need for accurate water level forecasting to support timely early warning and disaster response. Such needs are particularly acute in ungauged or under-monitored regions, where monitoring infrastructure remains insufficient. Unlike streamflow, water level directly reflects overtopping risk and thus provides more actionable information for disaster prevention systems.

1.2. National context: observation system gaps in South Korea

Despite the critical role of real-time water level data in flood forecasting, South Korea exhibits substantial disparities in the spatial distribution of gauging stations. As illustrated in Figure 1, over 95% of rivers are classified as regional rivers or small streams, yet monitoring infrastructure remains heavily concentrated along national rivers [5]. National rivers have an average of 4.25 stations per river, equivalent to 8.6 stations per 100 km of river length, whereas regional rivers and small streams are monitored by only 0.10 and 0.02 stations per river, respectively, corresponding to 1.4 and 0.8 stations per 100 km [6-7]. This imbalance significantly constrains the applicability of conventional data-driven models, while

Corresponding author email address: hongjo@yonsei.ac.kr

infrastructure expansion remains economically impractical. Consequently, there is an urgent need for forecasting methods specifically tailored to under-monitored and ungauged river systems, particularly those most vulnerable to flood-related impacts.

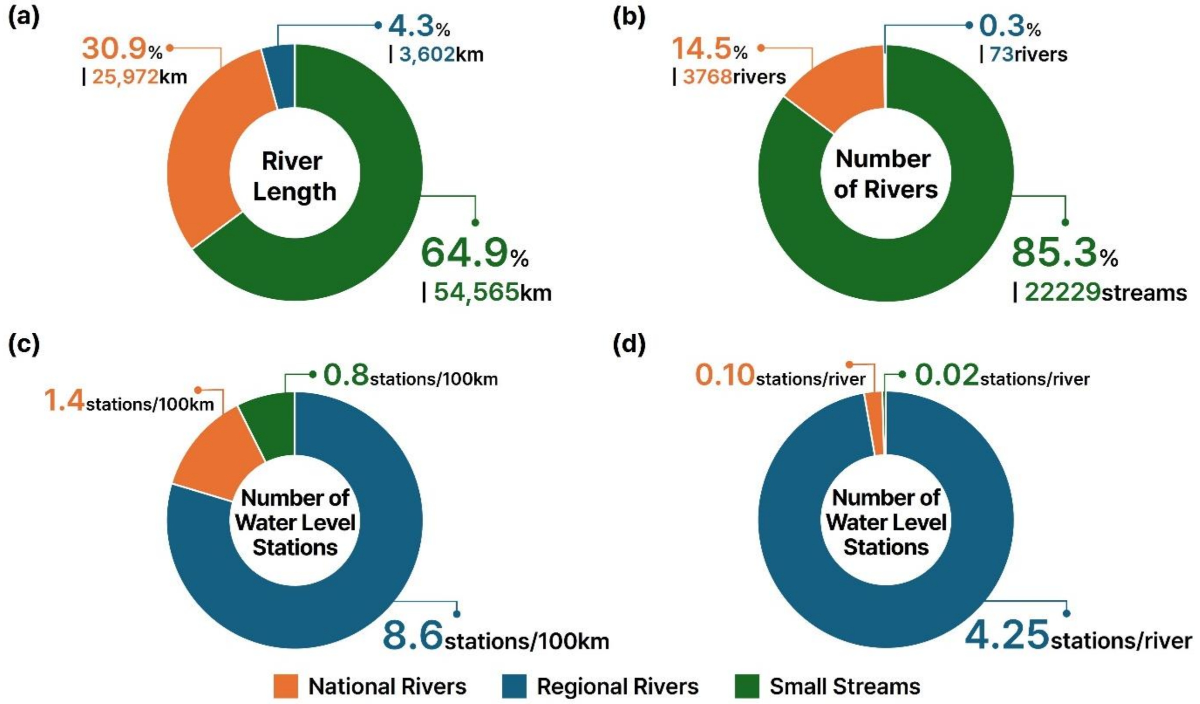

Fig. 1. (a) River length by classification (% and km); (b) Number of rivers by classification (% and count); (c) Number of water level stations per 100km of river by classification; (d) Number of water level stations per river by classification.

1.3. Previous approaches

Water level forecasting methods are generally classified into physics-based and data-driven approaches. Physics-based models, particularly hydrodynamic simulations, provide physically interpretable outputs but require intensive computation and high-quality input data, thereby limiting their scalability for real-time applications [8-11]. In ungauged or data-scarce basins, their construction and calibration become particularly challenging [12]. In contrast, data-driven models offer greater flexibility and do not require detailed prior knowledge of catchment characteristics [13-14]. However, their performance often deteriorates in regions with sparse historical data, increasing the risk of overfitting [9]. While hybrid models that combine physics-based and data-driven approaches can improve accuracy, their inherent complexity and reliance on high-quality inputs similarly constrain their use in real-time forecasting [9, 11]. Moreover, most deep learning studies have focused on specific gauged locations [15] or on upstream-downstream relationships using complete time series data [9, 13-14, 16-18]. Although regionalization and transfer learning have been applied to streamflow prediction in ungauged basins, these methods primarily target discharge or runoff rather than water level, the most direct indicator of overtopping risk [19-23]. While a few studies have investigated rainfall-event-specific training [18], research on event-centered, short-term water level forecasting in ungauged regions remains sparse.

1.4. Objectives and contributions

Despite recent advancements in water level forecasting, critical limitations remain, particularly in ungauged basins. Most existing models depend on long-term data from gauged sites and often underperform during extreme events or noisy conditions, thereby limiting their real-time applicability in ungauged basins. To address these challenges, we propose DeepCreek, a lightweight deep learning model designed to (1) operate effectively in data-scarce environments, (2) improve prediction accuracy through event-centered training, (3) ensure spatial generalization, and (4) maintain robustness under extreme and anomalous conditions.

The main contributions of this study are as follows:

- Development of a nationwide hourly hydrometeorological dataset from 232 sites in South Korea.
- Adaptation and enhancement of the TSMixer architecture to better capture temporal dynamics and inter-feature relationships critical for flood prediction.
- Implementation of a spatial cross-validation strategy to realistically simulate ungauged conditions.
- Demonstration of generalization and robustness through nationwide extreme-event experiments.

These contributions collectively establish a scalable foundation for real-time water level forecasting in regional rivers, where monitoring infrastructure is limited yet flood risks are substantial.

2. Methodology

2.1. Study area and data preprocessing

This study focuses on rivers and stream networks across South Korea. Hydrological and meteorological data, including rainfall (mm), water level (EL.m), air temperature (°C), relative humidity (%), dew point temperature (°C), sea-level pressure (hPa), wind speed (m/s), and wind direction (categorical compass values), were collected from the Han River Flood Control Office. These variables were selected based on prior studies emphasizing their relevance to hydrological modeling for ungauged basins [23]. A total of 232 observation sites were used to construct a nationwide, hourly time series dataset spanning multiple years. To ensure spatial consistency, each water level station was assigned the nearest rainfall station within 5 km and the nearest meteorological station within 25 km. Duplicate timestamps were removed, and stations with unstable water level records, identified by a difference greater than 6 meters between their 300th highest and lowest observations, were excluded. Additional preprocessing steps were performed to align the dataset toward rainfall-driven hydrological dynamics. Scenario-based segmentation was applied by extracting 24-hour windows (12 hours before and after) surrounding rainfall events lasting more than two hours.

2.2. Model architecture

We employed TSMixer, a recently proposed architecture for multivariate time series forecasting [24]. While traditional MLPs lack the ability to model temporal dependencies, TSMixer separates temporal and feature interactions through two distinct mixing blocks, first capturing temporal patterns for each feature, and second performs inter-feature mixing to learn cross-variable relationships as illustrated in Figure 2. To adapt the model to our forecasting task, we propose DeepCreek, an enhanced version of TSMixer with two key modifications. First, we progressively expand feature dimensionality across mixer blocks to mitigate representational bottlenecks and enhance the model's capacity to capture complex patterns. This design facilitates richer intermediate representations and improves generalization under dynamic and noisy hydrological conditions. Second, we increase the hidden dimension of the feed-forward network layers, thereby enhancing the model's ability to learn nonlinear transformations. This modification supports deeper abstraction of temporal and spatial dependencies, leading to improved forecasting performance. These enhancements enable the model to maintain a higher-capacity latent space across layers, rather than being constrained by low dimensional hidden vectors, which is critical for capturing complex spatiotemporal dynamics under real-world noise. A comprehensive overview of the architectural modifications and dimensional expansions is provided in Figure 3.

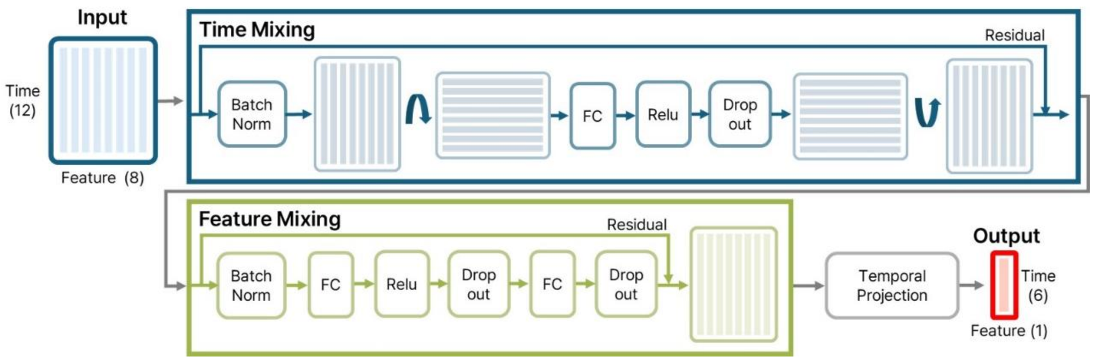

Fig. 2. Structure of TSMixer

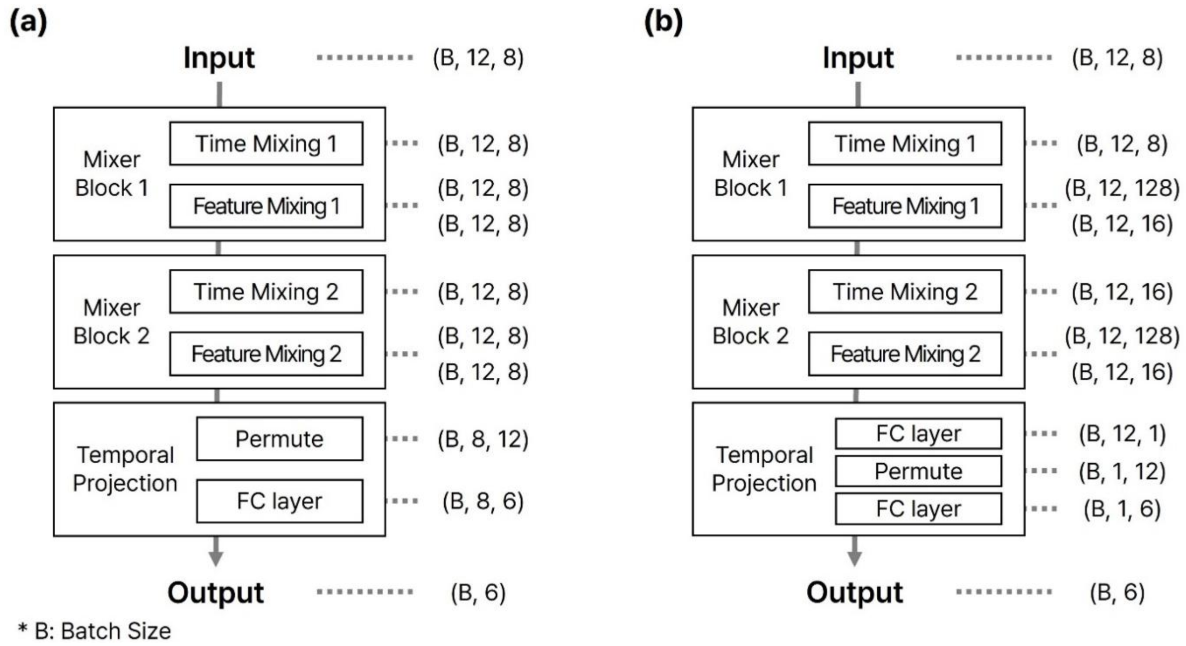

Fig. 3. Comparison of model architectures: (a) TSMixer; (b) DeepCreek.

2.3. Experimental setup and dataset composition

The dataset was spatially split into 80% training and 20% testing sites, ensuring complete spatial independence between train and test sets to realistically simulate forecasting in ungauged basins. In addition to the rainfall scenario-based segmentation described in Section 2.1, a general dataset of 18-hour sequences without missing values was also constructed. To evaluate model robustness under extreme rainfall conditions, a subset of the test set was extracted based on the World Meteorological Organization threshold, which defines rainfall exceeding 50 mm h^{-1} as an extreme event [25]. This Extreme-Scenario subset was evaluated separately to assess performance in high-risk flood situations. The detailed composition of all datasets is summarized in Table 1.

Table 1. Dataset composition for rainfall scenario-based, general, and extreme-scenario test sets.

Category	Rainfall Scenario-based			General sequences			Extreme-Scenario (Test Only)
	Train Set	Test Set	Total	Train Set	Test Set	Total	
No. of stations	185	47	232	185	47	232	36
No. of rainfall scenarios/sequences	91,488	23,748	115,236	31,773	10,564	42,337	170
No. of time steps	6,337,499	1,655,426	7,992,925	19,044,514	5,117,146	24,161,660	9,961

Each input consisted of eight hydrometeorological variables observed over the past 12 consecutive hours (t−12 to t−1): rainfall (rf), air temperature (ta), relative humidity (hm), dew point temperature (td), sea-level pressure (ps), wind_x, wind_y (decomposed wind speed and direction), and water level (wl). The model was trained to predict water levels for the next six hours (t to t+5), enabling short-term multi-step forecasting. A sliding window approach was adopted to generate input-output pairs at each valid time step.

All experiments were conducted on a Linux system (CPU: AMD Ryzen 7 5800X @ 4.85GHz, RAM: 32 GB, GPU: NVIDIA GeForce RTX 3090) using Python (3.10.16) and the PyTorch (2.6.0+cu124) deep learning framework. Model training was performed using the Adam optimizer with an initial learning rate of 0.001, a batch size of 8196, and mean squared error (MSE) as the loss function. Each model was trained for up to 100 epochs, with early stopping (patience = 5) and StepLR scheduler (decay factor =

0.5 every 10 epochs) to prevent overfitting and stabilize convergence. The same training configuration was consistently applied across all model variants to ensure fair comparison.

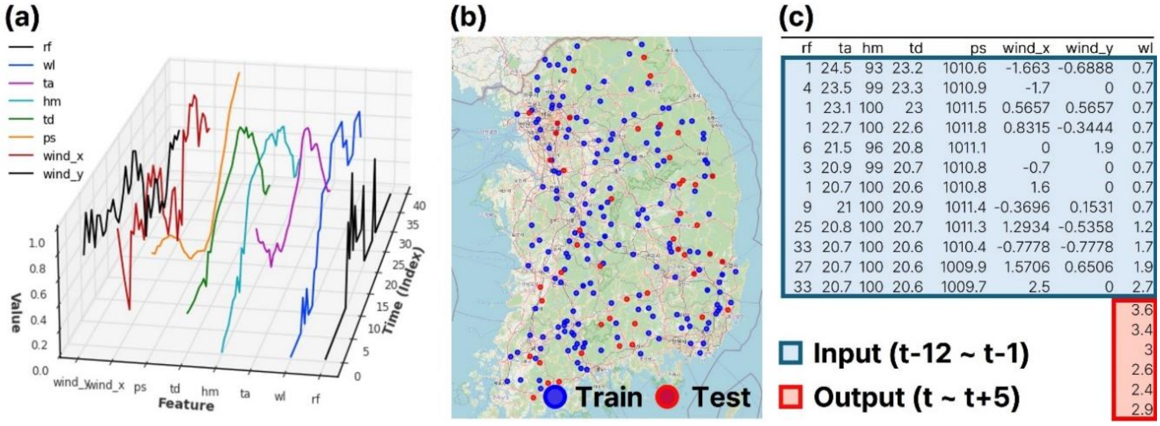

Fig. 4. (a) Example of normalized features over time; (b) Spatial distribution of train and test sites; (c) Input-output structure for one prediction instance.

2.4. Baseline models for comparison

To evaluate the performance of the proposed model, we conducted comparative experiments using widely adopted deep learning models that are commonly applied in recent water level forecasting studies. These include CNN [14], GRU [9, 15], LSTM [13, 15-18] and Bi-LSTM [13], which have been used either as standalone models or as components within hybrid or attention-based frameworks. All models were trained under the same experimental setup described in Section 2.3. Table 2 summarizes the architectural configurations of each baseline model.

Table 2. Architectural configurations of baseline models used for comparison.

Model	Type	Directionality	Key Architecture
CNN	CNN	-	2 Conv2D layers + MaxPool + 2 FC layers
GRU	RNN	Uni	2-layer GRU + 1 FC layer
LSTM	RNN	Uni	2-layer LSTM + 1 FC layer
Bi-LSTM	RNN	Bi	2-layer Bi-LSTM + 1 FC layer
TSMixer	MLP	-	Stacked MLPs for temporal and feature mixing
DeepCreek	MLP	-	TSMixer with output and hidden dimension expansion

3. Results

3.1. Evaluation metrics

To assess the predictive performance of the models, we adopted five widely used metrics, mean squared error (MSE), root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and the coefficient of determination (R^2). The metrics are defined as follows:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{1}$$

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

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

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \tag{4}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{5}$$

where y_i and $\hat{y}_i$ denote the observed and predicted water level at time i, $\bar{y}$ is the mean of the observed values, and n is the number of samples.

3.2. Performance comparison

Tables 3 and 4 compare the performance of DeepCreek and baseline models across all evaluation metrics under ungauged conditions. Table 3 presents results on rainfall scenario test set, while Table 4 extends the analysis to the extreme rainfall scenario test set, providing a robustness assessment under high-risk flood conditions. Across all settings, DeepCreek consistently outperformed CNN, GRU, LSTM, and Bi-LSTM, achieving significantly lower error rates and higher R^2 scores, particularly under extreme events. While models such as CNN and LSTM exhibited overfitting with marked performance degradation from training to test sets, DeepCreek maintained strong generalization across spatial and temporal variability.

Table 3. Model performance for water level prediction under ungauged conditions. (Scenario test set; trained on general sequences vs. event-centered sequences)

Model	Train type	MSE(m²)	RMSE(m)	MAE(m)	MAPE(%)	R^2
CNN	general	0.1714	0.4140	0.3348	44.0969	0.9977
GRU		11.2852	3.3593	0.3086	9.9442	0.8490
LSTM		11.5658	3.4009	0.3010	9.5703	0.8452
Bi-LSTM		11.2670	3.3566	0.3027	12.1142	0.8492
TSMixer		0.0418	0.2044	0.0757	6.2038	0.9994
DeepCreek (ours)		0.0280	0.1674	0.0744	9.8377	0.9996
CNN	scenario	0.2202	0.4692	0.2665	28.5412	0.9971
GRU		11.0610	3.3258	0.4497	40.9548	0.8520
LSTM		11.5123	3.3930	0.3428	14.4307	0.8459
Bi-LSTM		11.3556	3.3698	0.4156	28.7043	0.8480
TSMixer		0.0279	0.1670	0.0602	6.8828	0.9996
DeepCreek (ours)		**0.0267**	**0.1634**	**0.0594**	**6.5205**	**0.9996**

Table 4. Model performance for water level prediction under ungauged conditions. (Extreme scenario test set; trained on general vs. event-centered sequences)

Model	Train type	MSE(m²)	RMSE(m)	MAE(m)	MAPE(%)	R^2
CNN	general	1.3433	1.1590	0.5789	37.5395	0.9975
GRU		105.6633	10.2793	2.2401	18.0697	0.8032
LSTM		108.3523	10.4092	2.2823	19.3125	0.7982
Bi-LSTM		106.5497	10.3223	2.2289	19.9348	0.8016
TSMixer		1.3278	1.1523	0.3395	17.1734	0.9975
DeepCreek (ours)		1.2223	1.1056	0.2839	17.3792	0.9977
CNN	scenario	2.1563	1.4684	0.6323	30.6445	0.9960
GRU		103.3975	10.1685	2.4049	41.0707	0.8074
LSTM		107.7760	10.3815	2.3068	22.2475	0.7993
Bi-LSTM		106.0554	10.2983	2.3360	30.3467	0.8025
TSMixer		1.4340	1.1975	0.3012	18.2150	0.9973
DeepCreek (ours)		**1.2072**	**1.0987**	**0.2760**	**15.4311**	**0.9978**

3.3. Case analysis of extreme scenarios

Although the overall performance gap between DeepCreek and TSMixer was modest, substantial improvements emerged under specific extreme conditions. Table 5 presents the top 10 extreme test cases where DeepCreek achieved the largest gains, with RMSE over 95%, MAE above 90%. To further illustrate these differences, Figure 5 highlights four representative cases. Cases (a) and (b) correspond to actual extreme rainfall events, marked by abrupt and intense rainfall causing rapid water level changes. In these instances, DeepCreek successfully tracked the dynamics with stability, whereas TSMixer exhibited overshooting and instability. Cases (c) and (d) involve sequences with anomalous rainfall inputs exceeding 1,000 mm, likely due to sensor errors or extreme outliers. Despite these irregularities, DeepCreek maintained robust performance, while TSMixer experienced significant degradation. These findings suggest that DeepCreek not only excels during real extreme events but also demonstrates resilience to severe data anomalies—an essential attribute for flood forecasting in uncertain and data-scarce environments.

Table 5. Top 10 performance gains by DeepCreek over TSMixer under extreme scenarios. (unit: relative improvement [%])

Station name	MSE	RMSE	MAE	MAPE	R^2
Han River, Inje (Eoduwon Bridge)	99.88	96.48	95.00	95.48	100.06
Han River, Hoengseong (Anheung Bridge)	99.81	95.70	93.03	93.08	99.89
Han River, Huyeong Bridge	78.65	55.35	36.89	45.62	195.21
Seomjin River, Boseong (Boseonggang Dam)	76.98	52.74	57.49	57.48	98.88
Geum River, Daejeon (Wanchon Bridge)	38.96	22.98	21.64	25.18	219.98
Han River, Nayangju (Paldang Dam)	72.70	50.43	55.36	55.41	67.78
Nakdong River, Hamyang (Uitan-ri)	36.22	21.09	19.60	19.23	141.19
Han River, Seoul (Daegok Bridge)	36.80	21.06	19.40	21.18	55.88
Taehwa River, Ulsan (Taehwa Bridge)	26.11	15.11	15.19	14.34	31.27
Geum River, Geumsan (Munam Bridge)	18.55	10.14	21.00	23.54	22.88

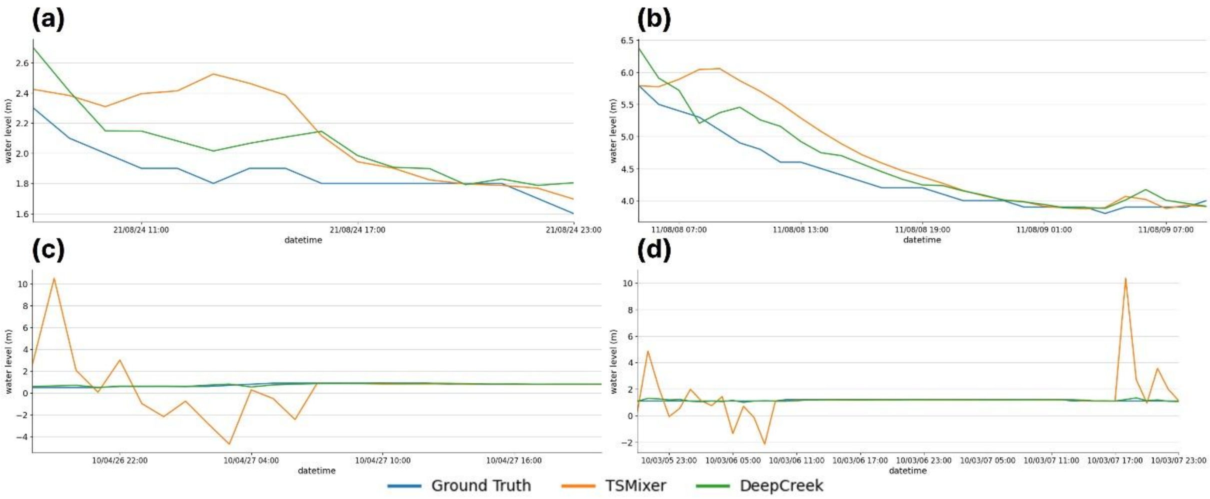

Fig. 5. Case studies comparing DeepCreek and TSMixer under extreme rainfall and outlier conditions: (a) Taehwa Bridge; (b) Uitan-ri; (c) Anheung Bridge; (d) Eoduwon Bridge.

4. Discussion

The superior performance of DeepCreek, particularly under extreme rainfall conditions, stems from its architectural enhancements. Expanding both the feature dimensionality and hidden dimensions mitigated information loss across mixer blocks, thereby improving stability under noisy fluctuations and abrupt changes. DeepCreek's architecture, which decouples temporal and feature dependencies, outperformed sequential models such as LSTM and GRU, which often suffer from overfitting or limited adaptability in sparse or highly variable conditions. By effectively handling heterogeneous inputs, DeepCreek offers a critical advantage for flood forecasting in ungauged basins. Furthermore, the model exhibited stable performance across varying lead times (t+1 to t+6), supporting its scalability across spatially diverse locations which enhances its practical utility for real-time deployment. Notably, all models were trained and evaluated without explicit outlier removal. Despite this, DeepCreek consistently maintained predictive stability across noisy real-world datasets, underscoring its robustness for operational use where extensive pre-cleaning is often impractical.

These promising results pave the way for further enhancements and broader applications, although several challenges remain. The scarcity of accessible data from small streams limited validation in truly ungauged basins, as the experiments were primarily conducted on national and regional rivers. Future work could enhance model performance by incorporating derived features, such as those extracted through frequency-domain decomposition methods like Empirical Mode Decomposition (EMD), to further enhance model adaptability to highly hydrological dynamics [16, 26].

5. Conclusion

This study introduced DeepCreek, a lightweight deep learning framework for short-term water level forecasting at a nationwide scale, specifically designed for ungauged and under-monitored river systems. By leveraging rainfall event-centered inputs and employing a spatially segregated training-testing strategy across 232 stations in South Korea, the model effectively simulated real-world operational scenarios in data-scarce regions. DeepCreek demonstrated robust predictive performance under diverse hydrological conditions, including extreme rainfall events and measurement anomalies, while maintaining scalability through the use of commonly available hydrometeorological variables. Its architecture, which decouples temporal and feature interactions, led to improved generalization and resilience compared to conventional sequential models. The architectural refinements, including the expansion of feature dimensionality across mixer blocks and the increase of hidden dimensions in the feed-forward network layers, contributed to enhanced model stability under noisy and abrupt hydrological conditions. These findings highlight the potential of event-driven, lightweight architectures in advancing real-time flood forecasting under increasing climate uncertainties. DeepCreek offers a practical and scalable solution for early warning systems, enabling reliable predictions even in regions with sparse monitoring infrastructure, and contributing to more equitable, adaptive, and sustainable disaster risk reduction strategies.

Acknowledgements

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT)(RS-2024-00457308). ※ MSIT: Ministry of Science and ICT

References

[1] U.S. Army Corps of Engineers (USACE), “National Levee Safety Guidelines.” Apr. 2024. Accessed: Apr. 16, 2025. [Online]. Available: https://www.leveesafety.org/pages/nlsg

[2] J. Danka and L. M. Zhang, “Dike Failure Mechanisms and Breaching Parameters,” *J. Geotech. Geoenvironmental Eng.*, vol. 141, no. 9, p. 04015039, Sep. 2015, doi: 10.1061/(ASCE)GT.1943-5606.0001335.

[3] K. Yun, B. Yoon, S. Lee, G. Yeo, and J. Hwangbo, “Urban Flood Disaster Management Technology, Supplementary Volume 8: Levee Breach and Floodplain Inundation Experiments and Analysis (First-Year Interim Report),” Accessed: Apr. 16, 2025. [Online]. Available: https://www.codil.or.kr/

[4] J. Kim, D. Lee, B. Choi, K. Kang, S. Cho, and J. Kim, “A Study on Levee Construction Techniques Using Riverbed Materials.” Accessed: Apr. 16, 2025. [Online]. Available: https://www.kaia.re.kr/

[5] J. Kim, “Legislative and Policy Tasks for Strengthening Small Stream Management,” Feb. 2023. Accessed: Apr. 17, 2025. [Online]. Available: https://www.nars.go.kr/

[6] Ministry of the Interior and Safety (MOIS), “Strengthening Small Stream Management in Response to Climate Change,” Dec. 28, 2023. Accessed: Apr. 17, 2025. [Online]. Available: https://www.mois.go.kr/

[7] Newsis, “Local rivers lack flood monitoring infrastructure, with limited water level stations installed,” *Newsis*, Oct. 14, 2024. Accessed: Apr. 17, 2025. [Online]. Available: https://www.newsis.com/

[8] F. Karim, M. A. Armin, D. Ahmedt-Aristizabal, L. Tychsen-Smith, and L. Petersson, “A Review of Hydrodynamic and Machine Learning Approaches for Flood Inundation Modeling,” *Water*, vol. 15, no. 3, Art. no. 3, Jan. 2023, doi: 10.3390/w15030566.

[9] Z. Dong *et al.*, “Superior performance of hybrid model in ungauged basins for real-time hourly water level forecasting – A case study on the Lancang-Mekong mainstream,” *J. Hydrol.*, vol. 633, p. 130941, Apr. 2024, doi: 10.1016/j.jhydrol.2024.130941.

[10] J. Teng, A. J. Jakeman, J. Vaze, B. F. W. Croke, D. Dutta, and S. Kim, “Flood inundation modelling: A review of methods, recent advances and uncertainty analysis,” *Environ. Model. Softw.*, vol. 90, pp. 201–216, Apr. 2017, doi: 10.1016/j.envsoft.2017.01.006.

[11] H. Zhai, M. Pan, and E. F. Wood, “Coupling Machine Learning Into Hydrodynamic Models to Improve River Modeling With Complex Boundary Conditions,” *Water Resour. Res.*, vol. 58, no. 12, p. e2022WR032183, Dec. 2022, doi: 10.1029/2022WR032183.

[12] R. B. Mudashiru, N. Sabtu, I. Abustan, and W. Balogun, “Flood hazard mapping methods: A review,” *J. Hydrol.*, vol. 603, p. 126846, Dec. 2021, doi: 10.1016/j.jhydrol.2021.126846.

[13] G. Li, Z. Shu, M. Lin, J. Zhang, X. Yan, and Z. Liu, "Comparison of strategies for multistep-ahead lake water level forecasting using deep learning models," *J. Clean. Prod.*, vol. 444, p. 141228, Mar. 2024, doi: 10.1016/j.jclepro.2024.141228.

[14] X. Ji *et al.*, "Spatio-temporal multivariable time vario-zoom network for water level forecasting based on high-resolution hydrological dataset," *J. Hydrol.*, vol. 634, p. 131060, May 2024, doi: 10.1016/j.jhydrol.2024.131060.

[15] W. Almikaeel, A. Šoltész, L. Čubanová, and D. Baroková, "Hydro-informer: a deep learning model for accurate water level and flood predictions," *Nat. Hazards*, Oct. 2024, doi: 10.1007/s11069-024-06949-8.

[16] D. Ren, Q. Hu, and T. Zhang, "EKLT: Kolmogorov-Arnold attention-driven LSTM with Transformer model for river water level prediction," *J. Hydrol.*, vol. 649, p. 132430, Mar. 2025, doi: 10.1016/j.jhydrol.2024.132430.

[17] G. Li *et al.*, "A combined hydrodynamic model and deep learning method to predict water level in ungauged rivers," *J. Hydrol.*, vol. 625, p. 130025, Oct. 2023, doi: 10.1016/j.jhydrol.2023.130025.

[18] D. Kim, J. Lee, J. Kim, M. Lee, W. Wang, and H. S. Kim, "Comparative analysis of long short-term memory and storage function model for flood water level forecasting of Bokha stream in NamHan River, Korea," *J. Hydrol.*, vol. 606, p. 127415, Mar. 2022, doi: 10.1016/j.jhydrol.2021.127415.

[19] R. Arsenault, Breton-Dufour ,Mélissa, Poulin ,Annie, Dallaire ,Gabrielle, and R. and Romero-Lopez, "Streamflow prediction in ungauged basins: analysis of regionalization methods in a hydrologically heterogeneous region of Mexico," *Hydrol. Sci. J.*, vol. 64, no. 11, pp. 1297–1311, Aug. 2019, doi: 10.1080/02626667.2019.1639716.

[20] P. Istalkar and B. Biswal, "Streamflow prediction in ungauged basins: How dissimilar are drainage basins?," *J. Hydrol.*, vol. 637, p. 131357, Jun. 2024, doi: 10.1016/j.jhydrol.2024.131357.

[21] Y. Xu *et al.*, "Deep transfer learning based on transformer for flood forecasting in data-sparse basins," *J. Hydrol.*, vol. 625, p. 129956, Oct. 2023, doi: 10.1016/j.jhydrol.2023.129956.

[22] R. Dasgupta, S. Das, G. Banerjee, and A. Mazumdar, "Revisit hydrological modeling in ungauged catchments comparing regionalization, satellite observations, and machine learning approaches," *HydroResearch*, vol. 7, pp. 15–31, Jan. 2024, doi: 10.1016/j.hydres.2023.11.001.

[23] J. Choi, J. Lee, and S. Kim, "Utilization of the Long Short-Term Memory network for predicting streamflow in ungauged basins in Korea," *Ecol. Eng.*, vol. 182, p. 106699, Sep. 2022, doi: 10.1016/j.ecoleng.2022.106699.

[24] S.-A. Chen, C.-L. Li, N. Yoder, S. O. Arik, and T. Pfister, "TSMixer: An All-MLP Architecture for Time Series Forecasting," Sep. 11, 2023, *arXiv*: arXiv:2303.06053. doi: 10.48550/arXiv.2303.06053.

[25] World Meteorological Organization (WMO), "Guide to Instruments and Methods of Observation," Geneva, Guide WMO-No.8, 2025. Accessed: Apr. 26, 2025. [Online]. Available: https://community.wmo.int/en/activity-areas/imop/wmo-no_8

[26] Y. Bai, W. Xing, L. Ding, Q. Yu, W. Song, and Y. Zhu, "Application of a Hybrid Model Based on Secondary Decomposition and ELM Neural Network in Water Level Prediction," *J. Hydrol. Eng.*, vol. 29, no. 2, p. 04024002, Apr. 2024, doi: 10.1061/JHYEFF.HEENG-5946.