Environmental Modelling and Software 197 (2026) 106832

Contents lists available at ScienceDirect

Environmental Modelling and Software

journal homepage: www.elsevier.com/locate/envsoft

FloodTransformer: Efficient real-time high-resolution flood forecasting

Zhanzhong Gu [a,1,*], Jiachen Kang [a,1,*], Wenzheng Jin [a], Feifei Tong [b], Y. Jay Guo [a], Wenjing Jia [a,*]

[a] School of Electrical and Data Engineering, University of Technology Sydney, NSW, 2007, Australia
[b] School of Architecture and Civil Engineering, The University of Adelaide, South Australia, 5005, Australia

ARTICLE INFO

Keywords:
Flood forecasting
Hydrodynamic simulation
Artificial intelligence
FloodTransformer

ABSTRACT

Flood forecasting is crucial for disaster planning and risk management, yet conventional hydrodynamic-based approaches are often slow in response and computationally intensive. We present a hybrid framework leveraging traditional hydrodynamic modelling with a novel AI model to enable accurate, real-time, and high-resolution flood prediction. To address the computational challenges of large-scale, dense flood prediction, we develop an efficient flood prediction model, *FloodTransformer*, which possesses three key novelties: variable-size cell embedding, tokenised time-sequence encoding, and physics-informed multi-task optimisation. These components effectively capture complex spatiotemporal dependencies, allowing accurate sequential predictions in a single run. Comprehensive evaluations on both simulated and historical flood events demonstrate FloodTransformer's excellent accuracy and efficiency: NSE 0.9445, KGE 0.9759 for water-depth prediction, and IoU 0.8180, F1 0.8997 for inundation classification, outperforming all comparative models. With 3s inference enabling multiple horizons in one pass, FloodTransformer offers a robust and practical solution for operational flood risk management.

1. Introduction

Floods are among the most destructive natural hazards globally, causing significant economic losses, infrastructure damage, and loss of human life every year (Tavakoli et al., 2025; Jonkman, 2005). As global climate change intensifies rainfall variability and increases the frequency of extreme events, the demand for accurate, real-time, and high-resolution flood forecasts is greater than ever (Byaruhanga et al., 2024; Hakim et al., 2024). Traditional flood forecasting systems, built on deterministic hydrodynamic models, have proven reliable in simulating flood behaviour under specific scenarios (Subbarayan et al., 2025; Al-Rawas et al., 2024; Xu and Gao, 2024; Samal et al., 2024). However, they are restricted by slow response time and expensive computational costs, difficult to scale for large catchments, and unable to provide real-time forecasting (Samantaray et al., 2023; Antwi-Agyakwa et al., 2023; Adnan et al., 2023). These challenges highlight the need for hybrid approaches that retain physical realism while enabling accurate and real-time prediction capabilities.

Recent advancements in artificial intelligence (AI) and data-driven modelling offer promising directions for enhancing flood forecasting (Hayder et al., 2023; Schumann et al., 2023). Deep learning, in particular, has shown promise in capturing spatiotemporal dynamics of floods from large-scale simulation and observational data (Zhang et al., 2023; Shao et al., 2024; Biswakalyani et al., 2024; Samantaray and Sahoo, 2024b; Samantaray et al., 2025). Several prior works have explored data-driven flood forecasting using neural networks. For instance, Convolutional Neural Network (CNN) based approaches have been effective in extracting spatial features from inundation state maps (Ouma and Omai, 2023; Guo et al., 2022), while Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) have been applied for temporal sequence prediction of water depths (Chen et al., 2023; Hu et al., 2025; Situ et al., 2024). However, such methods often fall short in handling high-dimensional spatiotemporal inputs.

The Transformer architecture (Vaswani et al., 2017), originally developed for natural language processing (NLP), has recently been adapted for spatiotemporal prediction tasks in weather and hydrology (Jamali and Mahdianpari, 2022; Castangia et al., 2023; Li et al., 2024; Zhou et al., 2025). Its self-attention mechanism allows it to capture long-range dependencies across time and space more efficiently than traditional deep learning models. However, many AI-based flood forecasting models struggle with the computational costs of high-resolution rasters across large terrain (Zhang et al., 2024; Al-Rawas et al., 2024). Additionally, there remains a trade-off in predictive

* Corresponding authors.
E-mail addresses: Zhanzhong.Gu@uts.edu.au (Z. Gu), Jiachen.Kang-1@uts.edu.au (J. Kang), Wenjing.Jia@uts.edu.au (W. Jia).
[1] Equal contribution.

https://doi.org/10.1016/j.envsoft.2025.106832
Received 28 August 2025; Received in revised form 31 October 2025; Accepted 9 December 2025
Available online 11 December 2025

focus—most methods focus on either classifying inundation state or estimating water depth, but seldom achieve both simultaneously at a high accuracy (Antwi-Agyakwa et al., 2023; Hakim et al., 2024). Finally, the lack of rasterised ground-truth data is prevalent across nearly all regions.

Latest advances highlight the growing adoption of hybrid AI-hydrodynamic frameworks in flood forecasting, combining physical consistency with data-driven efficiency. These approaches integrate AI models into traditional hydrodynamic modellings to improve predictive accuracy, reduce computational cost, and enhance spatial generalisability (Samantaray and Sahoo, 2024a; Huynh et al., 2025; Feng et al., 2025; Chand et al., 2025; Tran et al., 2025; Anik et al., 2025). Instead of relying on limited observation data, simulations generated by traditional hydrodynamic flood models are used for AI training. Despite these advances, most existing models still face challenges in heavy computational demands, cumulative errors in sequential predictions, and restricted real-time applicability, motivating the development of more efficient and transferable hybrid architectures.

To fill these significant research gaps, this study proposes an AI-hydrodynamic hybrid flood forecasting framework that integrates fine-scale hydrodynamic-based simulations with advanced deep learning techniques to achieve rapid, high-resolution (up to 5 m × 5 m), high-accuracy flood predictions. Our approach first leverages a widely used 3Di hydrodynamic flood model (3Di, 2025; Dahm et al., 2014; Liao and Ho, 2024; Chang et al., 2023) to simulate a comprehensive set of design rainfall scenarios derived from the Australian Rainfall and Runoff (ARR) guidelines (Ball et al., 2016). After rigorous validation and calibration against real-world gauge data and flood extent observations, hundreds of simulations under various initial water levels, boundary conditions, and rainfall durations, reflecting a wide range of flood events (Wang et al., 2023; Zhou et al., 2023b), are generated to serve as the training dataset for the AI model.

Predicting rasterised flooding over a large catchment and longer time horizon at a high spatial resolution requires computing the water depth for millions of cells at every time step, which is an extremely memory-intensive process that constrains the scalability of AI models. To address the computational challenges brought by high resolution and wide geographical coverage, we propose FloodTransformer, an efficient Transformer-based architecture designed for real-time, sequential flood forecasting. The model first leverages an innovative size and position embedding to extract spatial information from variable-sized grid cells, followed by a flood-specific Transformer block with a tokenised time-sequence encoding to capture the long-term temporal dependencies. This architecture significantly reduces computation overhead and achieves real-time multi-step predictions in one run. To ensure physically consistent results, the model incorporates physics-informed multi-task optimisation with integrated regulations, enhancing both the accuracy and hydrodynamic consistency of the flood forecasting.

In this paper, Wagga Wagga catchment, located in New South Wales (NSW), Australia, was selected as the study area due to its high flood susceptibility and global representativeness of urban flood challenges (NSW State Emergency Service, 2022; Bureau of Meteorology, 2024). Extensive simulations, calibrations, and evaluations have been conducted to evaluate the accuracy of predicted water depth rasters and the inundation state, using widely adopted hydrographic metrics and standard classification metrics, respectively. The results on both simulated datasets and historical flood events demonstrate that our proposed FloodTransformer achieves both high accuracy and computational efficiency.

The novelty of this study lies in three key areas:

- This study proposes an AI-hydrodynamic hybrid approach that integrates advanced deep learning models with physics-based hydrodynamic simulations, enabling real-time and reliable flood forecasting while preserving the physical consistency and interpretability of traditional hydrodynamic models.
- We design an efficient FloodTransformer model with novel variable-sized cell embeddings, which allows fast and effective processing of high-resolution multi-modality inputs for accurate flood forecasting across the large-scale catchment and urban regions.
- We propose an innovative tokenised time-sequence encoding and decoding architecture that captures long-term temporal dependencies with minimised computational costs and enables sequential predictions in a single run.

The remainder of this paper is structured as follows: Section 2 reviews related work in both conventional flood forecasting methods and advanced AI-based models. Section 3 details the flood modelling and simulation settings, including study area selection, schematisation, modelling, validation, calibration, and batch simulations. Section 4 explains in detail the architecture and training strategy of our efficient FloodTransformer model. Section 5 presents the experimental results and evaluation on both simulated datasets and historical flood events. Section 6 summarises principal findings and potential limitations. Finally, Section 7 concludes our work with discussions over future research directions.

2. Related works

2.1. Conventional flood forecasting methods

Traditional flood forecasting primarily relies on hydrological and hydrodynamic models such as SWAT (Rajib et al., 2020), HEC-RAS (Zhong et al., 2024; Do Lago et al., 2023; Lin et al., 2022), LISFLOOD (Rajib et al., 2020), and MIKE FLOOD (Young et al., 2025), which simulate rainfall–runoff dynamics based on physical principles. These models use topographic data like digital elevation models (DEMs), land cover, soil type, and precipitation to generate discharge or inundation maps. For instance, SWAT and LISFLOOD are widely used in continental flood simulation systems such as the European Flood Awareness System (EFAS) (Copernicus Emergency Management Service, 2024a) and Global Flood Awareness System (GLOFAS) (Copernicus Emergency Management Service, 2024b). These systems offer operational reliability in well-gauged regions but require high computational power and extensive calibrations (Young et al., 2025). Moreover, traditional simulation models, such as flood models built with HEC-RAS and MIKE FLOOD for reasonably sized catchments, require several hours to complete a single run, which hinders real-time applicability and limits their effectiveness for timely risk management and emergency planning (Rajib et al., 2020).

Some studies have adopted hybrid designs combining numerical simulations with real-time observations or remote sensing data for flood forecasting. For example, Xu and Gao (2024) and Zhong et al. (2024) used flood models driven by meteorological predictions and adjusted using observed inundation data. Others employed clustering or environment-based modelling to reduce simulation complexity (Wang et al., 2023). However, even such hybrid models remain sensitive to input uncertainty and often lack scalability.

Overall, the conventional hydrodynamic flood forecasting methods struggle with spatiotemporal flexibility, computational cost, and slow response in real-world scenarios, motivating the transition to AI-driven models.

2.2. AI-based flood forecasting models

AI-based approaches are increasingly used for flood prediction due to their capability to learn complex temporal and spatial patterns directly from data. Early models like Support Vector Machines (SVMs) (Hearst et al., 1998), Random Forests (Breiman, 2001), and shallow neural networks focused on water level or discharge prediction using rainfall and runoff time series (Samantaray et al., 2023; Hayder

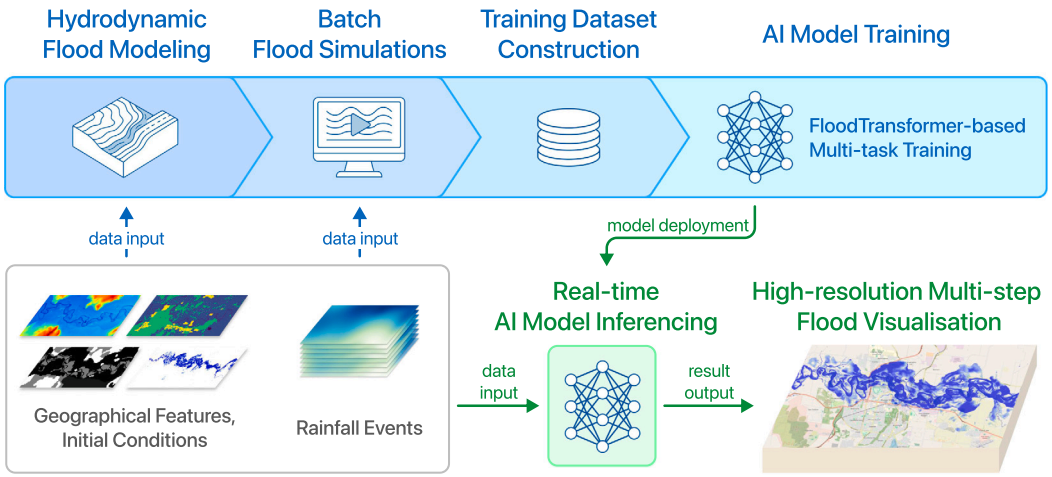


Fig. 1. The overall view of our AI-hydrodynamic hybrid approach. The top row illustrates the integration of multi-source data, including geographical and rainfall data, for hydrodynamic flood modelling and simulation. These simulations serve as the training dataset for the AI model, FloodTransformer. The bottom row shows the inference process, where the FloodTransformer processes multimodal inputs to generate high-resolution, multi-step predictions of flood inundation and water depth.

et al., 2023). These models achieved reasonable accuracy but lacked the ability to capture 2D dense spatiotemporal dynamics and were highly dependent on historical data from specific gauged sites (Elsadek et al., 2024; Gharakhanlou and Perez, 2023).

Deep learning methods, such as LSTMs and RNNs, have significantly improved temporal learning and prediction in time-series flood forecasting. LSTM-based models have demonstrated better performance in short-term flood forecasting and outperform traditional machine learning models in urban or riverine scenarios (Chen et al., 2023; Hu et al., 2025; Shao et al., 2024; Zhang et al., 2023; Samantaray et al., 2025). However, their effectiveness diminishes in high-resolution forecasting for large ungauged areas due to data scarcity and the need for extensive domain-specific tuning (Wajid et al., 2024; Al-Rawas et al., 2024). Some recent works begin to incorporate simulation-derived flood maps into AI pipelines (Adnan et al., 2023; Chowdhury et al., 2023; Dang et al., 2025), yet their designs still rely heavily on location-specific tuning or multimodal data fusion, limiting their real-world application in large catchments.

Later, spatiotemporally-aware models, such as 3D CNNs and CNN-RNNs, have been introduced to address these limitations. Studies show that integrating spatial inputs, *e.g.*, 2D rainfall grids, with temporal deep neural encodings can improve rasterised flood time-series prediction (Zhou et al., 2023a; Situ et al., 2024; Gao et al., 2024; Mahakur et al., 2025; Alkaabi et al., 2025). Recent attention-based and Transformer-based models have further enhanced spatiotemporal feature extraction over large basins (Jamali and Mahdianpari, 2022; Li et al., 2024; Zhou et al., 2025), advancing global flood forecasting efforts using geospatial and meteorological datasets, typically at resolutions of tens of kilometres (Zhang et al., 2024; Nearing et al., 2024). However, these models face significant computational challenges when applied to large-scale, high-resolution (*e.g.*, 5 m×5 m) flood forecasting. Furthermore, most approaches focus on either inundation state classification or water depth prediction, rarely achieving both with high accuracy (Antwi-Agyakwa et al., 2023; Hakim et al., 2024).

Hybrid frameworks combining hydrodynamic models with AI have recently gained attraction in flood forecasting. Huynh et al. (2025) proposed a distributed physics–AI model that learns corrections to internal hydrological fluxes while preserving conceptual structure, enhancing high-resolution flood modelling. Feng et al. (2025) compared physics-informed and data-driven models, offering an adaptive hybrid strategy to balance interpretability and predictive power. Chand et al. (2025) developed a CNN–GRU hybrid for hourly flood-index forecasting in island settings, achieving high short-term accuracy. Tran et al. (2025) presented a hybrid framework that substantially improves accuracy and real-time efficiency over baseline hydrodynamic models. Anik et al. (2025) reviewed urban-flood modelling trends, highlighting the shift towards AI-driven real-time systems. These studies show that hybrid models improve accuracy, reduce computational cost, and enhance generalisability, though challenges remain, including high computational cost, error aggregation for multiple predictions, and limited real-time efficiency.

To address these gaps, this work presents an efficient and effective FloodTransformer approach for real-time, large-scale, multi-step, and high-resolution flood forecasting (see Fig. 1). This model is trained on rigorously constructed hydrodynamic flood modelling and extensive simulations, with the following section explaining the detailed setup and processes, including flood study area selection, schematisation, modelling, validation, calibration, and batch simulation.

3. Flood modelling and simulation

Our AI-hydrodynamic approach starts with hydrodynamic-based flood modelling. We leverage a widely used hydrodynamic flood modelling software, 3Di (3Di, 2025), to model the mapped Wagga Wagga region and generate simulations for training, which involves several key steps: (1) study area selection and schematisation, which defines the boundary of the study area and its foundational spatial characters, such as catchment, meshing settings, and land use; (2) modelling, which constructs detailed topographic and hydrodynamic parameters; (3) calibration and validation with rating curves and flood inundation map; and (4) batch simulation under various initial conditions and rainfall patterns, detailed next.

3.1. Study area selection and schematisation

As shown in Fig. 2, in this work, we select Wagga Wagga, located in New South Wales (NSW), Australia, as the study area.

The region's combination of dense development, floodplain terrain, and dynamic river system provides a suitable testbed for evaluating real-time, high-resolution flood forecasting models applicable to similar settings worldwide. According to the NSW State Emergency Service (SES) (NSW State Emergency Service, 2022) and the Bureau of Meteorology (BoM) (Bureau of Meteorology, 2024) , the region is prone to intense rainfall events, which substantially increase the risk of riverine and flash flooding. Positioned along the Murrumbidgee River, Wagga Wagga has experienced numerous major flood events, including the

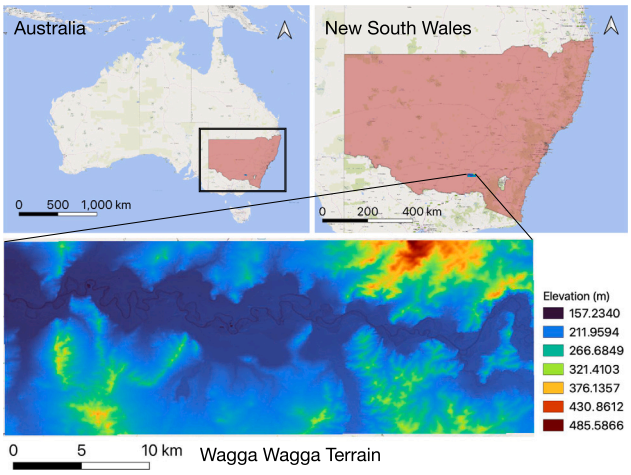


Fig. 2. Illustration of the geographic scope of our study area, the Wagga Wagga catchment in New South Wales, Australia.

devastating 2012 and 2022 floods that led to the evacuation of over 8000 residents and widespread infrastructure damage.[2,3] The city's combination of urban development, floodplain topography, and river dynamics presents a critical and applicable case for developing and validating real-time high-resolution flood forecasting approaches. Applying our proposed method in this area not only supports local disaster planning but also provides insights transferable to other flood-prone regions in Australia and beyond.

We developed a high-resolution 2D schematisation to represent the selected Wagga Wagga catchment, including detailed spatial and hydrodynamic features required by 3Di to generate the computational grid and sub-grid refinements. As shown in Fig. 3, the spatial extent and topography are illustrated, and a total of $346k$ variable-size cells with the finest resolution 5 m × 5 m are defined, the same as the DEM data, to represent the hydrodynamic network and terrain characteristics of our study area in detail.

3.2. Flood modelling

For flood modelling, we employ the flood modelling software 3Di (3Di, 2025)'s sub-grid technique to map the large area with the highest spatial resolution of 5 m × 5 m for the river channel, and the lowest resolution of 160 m × 160 m for areas where inundation is unlikely. In addition, multiple sets of data, including DEM, roughness, infiltration rate, land cover, land use, and boundary conditions, are sourced for flood modelling.

The DEM is a fundamental input for accurate flood modelling, representing the terrain surface that governs water flow paths and accumulation. In this work, the Geoscience Australia 5 Metre DEM (Geoscience Australia, 2015), derived from LiDAR surveys between 2001 and 2015, is used.

The roughness coefficient and infiltration rate are based on land use and soil type data. The land use data are sourced from the Australian Bureau of Agricultural and Resource Economics and Sciences (ABARES) (ABARES, 2025). It is a seamless raster dataset that combines land use data compiled at a resolution of 50 m × 50 m. Soil type data are sourced from the NSW Department of Climate Change, Energy, the Environment and Water's Great Soil Group (GSG) Soil Type map of NSW (NSW Department of Climate Change, Energy, the Environment and Water, 2021), which provides soil types across NSW, including over 55 datasets of multiple scales.

Land use features, particularly physical obstacles, play a critical role in flood modelling as they can significantly impact flow paths, water accumulations, and inundation extents. Among various types of obstacles, roads and levees are particularly important due to their function in controlling and redirecting floodwaters. In this study, we obtain the road data from the NSW Open Data Portal (Transport for NSW), which provides GIS-based road segments, including state, regional, and local roads (Transport for NSW, 2024). The levee data is from the NSW SES's levee study for emergency management, which provides detailed information on urban levees in NSW, including levee locations, design heights, and overtopping risks (Webb and Ltd, 2008). These attributes of key roads and levees are collected and schematised as obstacles in the 3Di modelling, ensuring accurate hydrodynamic influence in the simulation.

Boundary conditions define the interactions between the modelled area and its surroundings by specifying how water flows enter or exit the mapped area. In this work, horizontal 2D boundary condition lines are applied along the left (outflow) and right (inflow) edges of the schematisation area, representing the upstream and downstream flow of the Murrumbidgee River. All boundary conditions are designed to meet the technical requirements in 3Di modelling. Specifically, these boundary conditions only intersect one or more computational cells that are uniform in size and aligned horizontally, avoiding diagonal placement and grid refinement in boundary areas. This ensures stability and accuracy in simulating flow interactions at the model edges.

3.3. Calibration and validation

We conduct comprehensive calibration and validation of the 3Di flood modelling to ensure its accuracy and reliability. We use the rating curve method (Campbell and Bauder, 1940; Ocio et al., 2017) for calibration and validation, which compares and aligns the simulated rating curves with observed rating curves from historical gauge records. Moreover, we validate the predicted inundation state by comparing it with historical flood maps (WMAwater, 2018).

Rating Curve: A rating curve defines the relationship between water level and discharge at a specific location along a river, typically obtained from long-term gauge records. By comparing the simulated water levels and discharges at the gauges with the observation records, the accuracy and reliability of our flood modelling can be assessed.

In this study, rating curves from existing river gauges across the Wagga Wagga catchment are used as the primary basis for hydrodynamic model calibration and verification. Observed discharge and water level records were obtained from the Australian Bureau of Meteorology (BoM),[4] which provides publicly accessible and quality-assured datasets derived from river monitoring networks. Corresponding rainfall records are collected from BoM's regional weather stations to ensure consistency between the hydrological inputs and observed river responses.

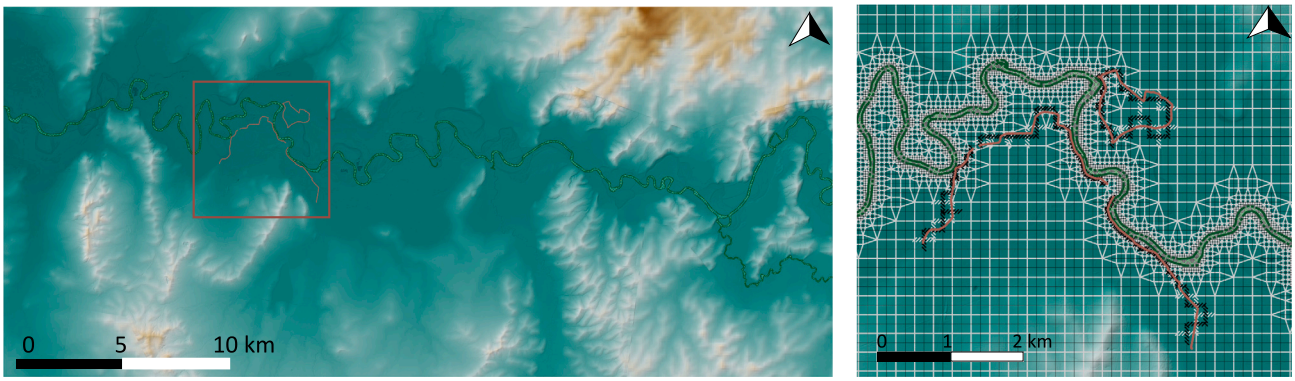
The 3Di model is configured using high-resolution digital elevation data, land-use maps, and boundaries to accurately represent the Wagga Wagga terrain (see more details in Section 3.2). The modelling settings, boundary conditions, and rainfall schemes followed standard 3Di simulation practices, with all configurations and parameters documented and stored within the 3Di online environment, making the whole process entirely reproducible and transparent.

For calibration, simulated discharge–water level relationships are extracted at each gauge location and plotted against the corresponding

[2] ABC News, "Floodwaters cut town of Forbes", 2012. Available at: https://www.abc.net.au/news/2012-03-07/hundreds-gear-up-for-another-in-evacuation-centres/3874974.

[3] ABC News, "Floodwaters have peaked at Forbes and Wagga Wagga in central NSW", 2022. Available at: https://www.abc.net.au/news/2022-11-06/floodwaters-have-peaked-at-forbes-and-wagga-wagga/101621508.

[4] https://www.bom.gov.au/waterdata/



(a) (b)

Fig. 3. Overview of the 3Di modelling scope and detailed meshes. Left: 3Di schematisation of the whole study area. Right: variable-size cells of the selected region within the red square shown on the left.

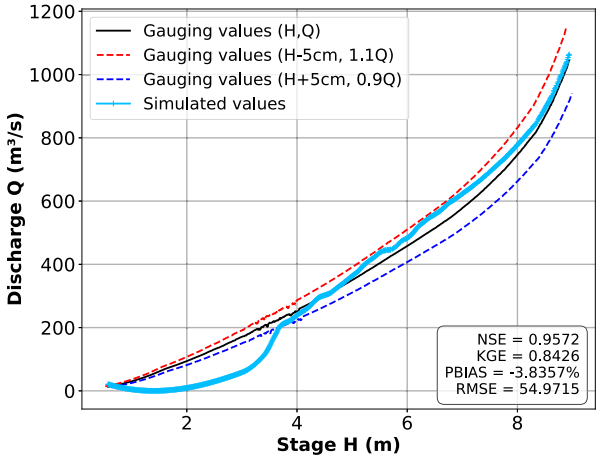


Fig. 4. The rating curve of the simulation data against the observed gauging values. Quantitative evaluation metrics are also reported, including NSE, KGE, PBIAS and RMSE.

observed rating curves. Iterative adjustments are applied to modelling and boundary parameters until optimal agreement is achieved across all gauges. Further validation is subsequently performed using independent flood events not included in the calibration phase, ensuring the model's robustness and generalisation across varying hydrodynamic conditions.

The rating curves shown in Fig. 4 illustrate the relationship between stage (H, in meters) and discharge (Q, in cubic meters per second) for both gauge readings and simulated values. The black curve represents the gauging records, while the two dashed curves (red and blue) define a tolerance band: a horizontal shift of ± 0.05 m in water depth ($H \pm 0.05$) and a vertical adjustment of $\pm 10\%$ in discharge ($Q \times 0.9$ to $Q \times 1.1$), which is commonly applied in hydrological practice (Lang et al., 2010). This band is due to measurement uncertainty and operational variability in gauging readings.

Our simulated curve (shown in sky blue) falls almost entirely within this tolerance envelope, confirming that the model reliably reproduces the stage–discharge relationship under typical flow conditions. This strong agreement indicates that our flood modelling accurately captures the hydrodynamic behaviour of the water flow during a wide range of flood events. It is noteworthy that slight deviations occur at very low stages ($H < 3$ m), which is consistent with previous studies reporting that empirical or physically based rating formulas often underperform in shallow-flow regimes (Lang et al., 2010). Such discrepancies at low stages are therefore expected and do not undermine the overall reliability of the model across operationally relevant ranges.

To quantitatively evaluate the model's accuracy and robustness, four widely used metrics in hydrology are adopted: the Nash–Sutcliffe Efficiency (NSE) (Gupta et al., 2009), the Kling–Gupta Efficiency (KGE) (Knoben et al., 2019), the Mean Absolute Error (MAE) (Willmott and Matsuura, 2005), and the Root Mean Square Error (RMSE) (Moriasi et al., 2015) (see Appendix B.1 for more details). As shown in Fig. 4, the NSE reaches 0.9572, indicating excellent agreement between simulated and observed discharge values. The KGE is 0.8426, reflecting strong performance across correlation, bias, and variability components. The PBIAS is –3.8357%, suggesting a slight underestimation of discharge but well within acceptable hydrological modelling thresholds. Additionally, the RMSE is 54.9715 m^3/s, demonstrating low overall deviation and confirming that the model maintains precision even under varying flow conditions. These results further prove the model's capability to reproduce stage–discharge dynamics with high fidelity and operational reliability.

Flood Inundation Map: We further validate the performance of our 3Di flood simulations by comparing them with the flood inundation map from real-world historical flood events. Fig. 5 shows the inundation map of our simulations against that of the 2018 flood events on the Wagga Wagga area (WMAwater, 2018). The high degree of spatial overlap between the two demonstrates that our flood modelling is capable of accurately reproducing real-world flood events. The alignment is particularly strong along the main floodplain, where the majority of the flood occurred. While there are minor discrepancies at the edges, these are largely due to uncertainties in the observed extent map and limitations in input data, such as coarse rainfall resolution.

For evaluating the spatial agreement, we adopt four widely used spatial classification metrics: Intersection over Union (IoU) (Zeng et al., 2024), Precision, Recall, and F1 score (Razali et al., 2020) (see Appendix B.2 for more details). The results are highly encouraging, with an IoU of 0.8531, indicating that over 85% of the predicted inundation area overlaps with the observed flood extent. The Precision score of 0.8945 reflects a low false positive rate, confirming that most predicted flooded areas are indeed correct. The Recall score reaches 0.9331, suggesting that the model successfully captures the majority of the actual inundated regions. The F1 score, which balances Precision and Recall, is 0.9136, further validating the model's overall reliability.

3.4. Flood simulation

The calibrated flood model is then used to generate simulations for training. To generate a comprehensive set of flood scenarios, we design a wide range of simulation conditions by systematically setting key

Table 1
Statistical summary of key input variables for training, validation, and test datasets. The X-velocity and Y-velocity refer to flow velocity in the X and Y directions, respectively.

Split	Variable	Min	Mean	Max	Kurtosis	Skewness	Std
Training (120)	Rainfall (m^3/s)	0.0000e+00	3.9155e–04	4.9705e–03	1.0317e+01	3.3979e+00	1.1520e–03
	Elevation (m)	1.5583e+02	1.8133e+02	2.2149e+02	3.1633e+00	1.7862e+00	1.2810e+01
	X-velocity (m/s)	0.0000e+00	2.6914e–02	2.3849e–01	5.0046e+00	2.5047e+00	6.3440e–02
	Y-velocity (m/s)	0.0000e+00	1.1925e–02	6.2937e–01	1.2515e+00	1.5819e+00	1.8877e–01
	Water depth (m)	0.0000e+00	3.6859e–02	8.5781e–01	–8.6502e–01	4.9371e–01	2.4772e–02
Validation (15)	Rainfall (m^3/s)	0.0000e+00	2.0039e–04	2.4889e–03	9.8707e+00	3.1789e+00	6.0492e–04
	Elevation (m)	1.5583e+02	1.8133e+02	2.2149e+02	3.1633e+00	1.7862e+00	1.2810e+01
	X-velocity (m/s)	0.0000e+00	2.4314e–02	2.2524e–01	5.3536e+00	2.5786e+00	5.9473e–02
	Y-velocity (m/s)	0.0000e+00	1.1781e–02	6.3847e–01	1.3868e+00	1.6271e+00	1.9102e–01
	Water depth (m)	0.0000e+00	3.7691e–02	8.8285e–01	–7.8482e–01	5.2954e–01	2.5637e–02
Test (15)	Rainfall (m^3/s)	0.0000e+00	1.8858e–04	2.2111e–03	8.4364e+00	3.1034e+00	5.3265e–04
	Elevation (m)	1.5583e+02	1.8133e+02	2.2149e+02	3.1633e+00	1.7862e+00	1.2810e+01
	X-velocity (m/s)	0.0000e+00	2.6149e–02	2.3862e–01	5.2430e+00	2.5543e+00	6.3100e–02
	Y-velocity (m/s)	0.0000e+00	1.2550e–02	6.6927e–01	1.5106e+00	1.6063e+00	2.0110e–01
	Water depth (m)	0.0000e+00	4.0102e–02	9.4104e+00	–8.1321e–01	5.0543e–01	2.6859e–02

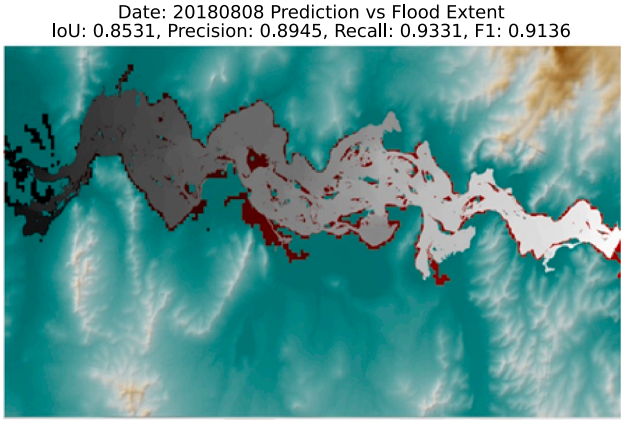


Fig. 5. The simulation results (red) against the 2018 flood extent map (grey) within the Wagga Wagga terrain. Quantitative evaluation metrics are also reported, including IoU, Precision, Recall, and F1 score.

simulation parameters as summarised in Table 2. The simulations adopt multiple combinations of initial water levels, boundary conditions, rainfall durations, simulation periods, and rainfall event probabilities. These parameters are selected based on ARR guidelines (Ball et al., 2016) and are intended to capture both typical and extreme flood events.

Initial Conditions: Three distinct initial water level conditions representing low, moderate, and high water levels are defined. These rasterised water level maps define the starting water depth throughout the study area and significantly influence how the simulation evolves, particularly in the early stages. Physically, these initial conditions represent different land saturation states before potential flood events. The initial conditions also include the startup boundary conditions, which define the upstream inflow and the downstream outflow at the boundary lines. Consistent with the initial water level, the boundary condition also has three levels, *i.e.*, low, middle, and high. This setting enhances the reliability of accurate flood modelling.

Rainfall Design: The rainfall inputs for the simulations are designed following the ARR guidelines, which define rainfall intensities, durations, and frequencies across the study area. As shown in Table 2, the simulation covers a wide range of rainfall durations—from short, flash rainfalls (*e.g.*, 30 min) to prolonged events up to 4 days (5760 min), with varying Annual Exceedance Probabilities (AEPs) and Average Recurrence Intervals (ARIs) to reflect both frequent and rare events (Ball et al., 2016). For instance, common ARIs such as 1, 5, and 10 years are included alongside more extreme conditions like the 500- and 1000 year events.

Table 2
Simulation condition categories and their respective variates, including varying Annual Exceedance Probabilities (AEPs) and Average Recurrence Intervals (ARIs) to reflect a wide range of rainfall events, from frequent to rare occurrences.

Conditions	Variates
Initial water level (m)	low, middle, high
Boundary flow (m^3/s)	low, middle, high
Rainfall (minutes)	60, 180, 360, 720, 1440, 2880, 5760
Simulation (hours)	12, 24, 48, 72, 120
AEP types (%)	63.2, 50, 20, 10, 5, 2, 1, 0.5, 0.2, 0.1
ARI types (years)	1, 2, 5, 10, 20, 50, 100, 200, 500, 1k

Batch Simulations: In light of the diverse simulation scenarios, we leverage the 3Di's API to automate the entire simulation workflow, enabling efficient batch processing across all defined conditions.

3.5. Dataset construction

To construct a comprehensive and representative dataset for AI model training and evaluation, a total of 150 hydrodynamic simulations are generated under diverse boundary and meteorological conditions shown in Table 2. These simulations are designed to capture a wide range of flood scenarios by systematically varying key factors: initial water levels, boundary flows, rainfall durations, simulation periods, and rainfall events represented by AEPs and ARIs.

To ensure balanced and representative subsets, we adopt a category-based splitting strategy rather than a simple random division. Simulations are grouped by key event attributes (ARI/AEPs, rainfall/durations, and initial water level/boundary flows) using the quantile binning method (Cox, 2018). As shown in Table 1, the simulations are split into training (80%), validation (10%), and test (10%) sets, corresponding to 120, 15, and 15 simulations, respectively. Simulations derived from the same rainfall event are kept together to avoid information leakage, ensuring that no simulation from a specific rainfall event appears in any other subset. This guarantees that the training, validation, and test sets are completely separate and independent, providing a robust evaluation of the model's generalisation performance.

Five key variables are listed in Table 1, including rainfall, elevation, X-velocity, Y-velocity, and water depth. Elevation remains consistent across all splits, while the other variables display noticeable skewness and kurtosis, reflecting the presence of extreme hydrological events. Rainfall and flow velocities generally have low mean values with considerable variability, whereas water depth shows moderate averages but includes occasional high extremes. These statistical characteristics highlight the dataset's diversity and complexity, making it suitable for robust AI model training and evaluation.

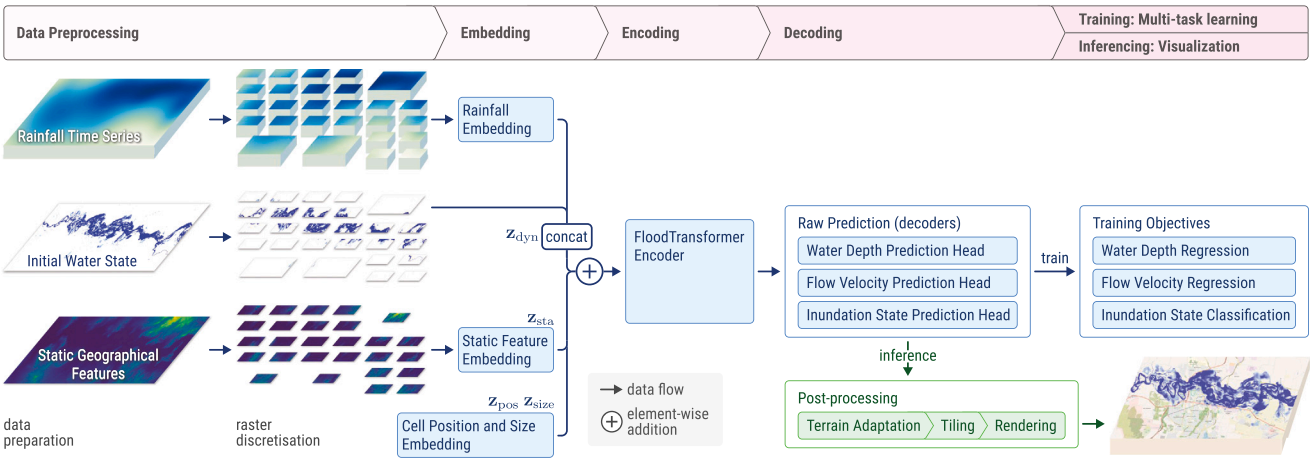


Fig. 6. Illustration of the proposed FloodTransformer model. FloodTransformer encodes three types of rasterised input data: initial water level, forecasted rainfall series, and static geographical features. It then incorporates novel position and size embeddings and dynamic feature embeddings to capture complex temporal and spatial dependencies across the input features. Multi-task optimisation with integrated regulations is adopted to predict both water depth and flood inundation maps with high accuracy.

Overall, the simulation dataset covers a wide range of flood patterns under diverse hydrological conditions, capturing both typical and extreme rainfall events. These comprehensive and high-quality simulations are both sufficient and well-suited to serve as the training dataset for our proposed FloodTransformer model, as detailed in the next section.

4. FloodTransformer

Based on the simulation results, we leverage a Vision Transformer (ViT) styled architecture (Dosovitskiy et al., 2021) and develop and train an efficient flood forecasting model, named FloodTransformer. To address the computational challenges associated with adopting ViT for large-scale, high-resolution rasterised spatiotemporal analysis, FloodTransformer directly encodes variable-sized grid cells and tokenises time-sequence embeddings to capture multi-modal temporal dependencies without significant computational overhead. This temporal tokenisation also ensures efficient sequential predictions in a single run, minimising error accumulation in iterative prediction approaches.

Formally, let the study area, $\mathcal{A}$, be discretised into a set of N computational cells according to the grid schematisation in the 3Di hydrodynamic modelling, indexed by $i = 1, \dots, N$. Our goal is to develop a deep learning model, denoted by $\mathcal{F}_\theta$, that is capable of learning a mapping, from a set of predictive variables, denoted as $\mathcal{X}$, to future flood states $\mathcal{Y}_{\text{sim}}$, describing the flood evolution for each cell i and for each future timestep $t \in \{1, \dots, T\}$, where T is the forecast horizon. In our AI-hydrodynamic hybrid approach, the ground-truth data $\mathcal{Y}_{\text{sim}}$ is generated from the hydrodynamic simulations. If we denote $\hat{\mathcal{Y}}$ as the model prediction, this mapping can be expressed as: $\hat{\mathcal{Y}} = \mathcal{F}_\theta(\mathcal{X})$. The ViT-style architecture of the FloodTransformer model $\mathcal{F}_\theta$ can proficiently capture long-range dependencies in intricate spatiotemporal sequences. The model parameters θ are optimised by minimising a composite loss function that evaluates the discrepancy between the model's predictions $\hat{\mathcal{Y}}$ and the ground truth $\mathcal{Y}_{\text{sim}}$. FloodTransformer operates in a non-autoregressive manner, which enables end-to-end, high-resolution, large-scale flood prediction over multiple time steps, all generated in a single forward pass. Fig. 6 shows the architecture of the proposed FloodTransformer model.

In the following sections, we provide details of the construction of the multimodal dataset (Section 4.1), variable-size cell representations (Section 4.2), and multi-task training with regulations (Section 4.3).

4.1. Multimodal dataset construction

The multimodal dataset for training and evaluating FloodTransformer is synthesised from two primary sources: (1) static geospatial data of the study area, which comprises multi-channel raster maps representing the DEM, surface roughness, and initial infiltration rate. (2) time-series rainfall forecast and time-series water levels from a series of 3Di hydrodynamic simulations, provided in NetCDF format (Rew and Davis, 1990), as described in Section 3.4. These raw data sources are processed to construct the input $\mathcal{X}$ and the ground truth labels $\mathcal{Y}_{\text{sim}}$.

The model inputs $\mathcal{X}$ consist of three components:

Static Geographical Features (S). To derive a rich representation for the static attributes of each cell, the DEM, roughness, and infiltration maps are first stacked to form a 3-channel geospatial image. For each cell i, we crop a square patch of $P \times P$ pixels from this composite image, centred at the cell's location. This patch, which captures the localised geographical context, is then processed by a pre-trained image encoder to produce a dense feature vector $\mathbf{s}_i \in \mathbb{R}^{D_s}$, and $S = \{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_N\}$.

Initial Water State ($\mathbf{d}^0, \mathbf{h}^0$). The 3Di model provides water levels, which we transform into water depths to enhance model stability. The water depth d_i^t for cell i at timestep t is calculated by subtracting the minimum DEM value within that cell from the water level and clipping any negative values to zero. The initial water depth vector (at timestep $t = 0$) can be denoted as $\mathbf{d}^0 = [d_1^0, d_2^0, \dots, d_N^0] \in \mathbb{R}^N$, The corresponding initial inundation state vector, $\mathbf{h}^0 = [h_1^0, h_2^0, \dots, h_N^0] \in \mathbb{R}^N$, is a binary mask derived from $\mathbf{d}^0$, where $h_i^0 = \mathbb{I}(d_i^0 > 0)$ and $\mathbb{I}(\cdot)$ is the indicator function.

Rainfall ($\mathcal{R}$). The rainfall data is directly extracted from the simulation files. This yields a matrix of rainfall intensities $\mathcal{R} = \{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N\}$ of shape $N \times T$, where each row vector $\mathbf{r}_i = [r_i^1, r_i^2, \dots, r_i^T]$ represents the rainfall for cell i over the prediction horizon of T timesteps.

The ground truth labels $\mathcal{Y}_{\text{sim}}$, used for model supervision, consist of the following three components:

Water Depth Variation (ΔD). To stabilise the training dynamics, the model is tasked with predicting the temporal change in water depth rather than its absolute value. The regression target for cell i at time t is therefore the difference $\Delta d_i^t = d_i^t - d_i^{t-1}$. For each training instance, these values are aggregated into a ground truth matrix $\Delta D \in \mathbb{R}^{N \times T}$, where $(\Delta D)_{i,t} = \Delta d_i^t$.

Inundation State ($\mathcal{H}$). Similar to $\mathbf{h}^0$, the inundation state h_i^t is defined as $h_i^t = \mathbb{I}(d_i^t > 0)$ for each cell i and timestep $t \in \{1, \dots, T\}$. These are elements in a binary classification target matrix $\mathcal{H} \in \mathbb{R}^{N \times T}$ in the dataset.

Flow Velocity ($\mathcal{U}, \mathcal{V}$). To further align the training objective with a physics-informed learning paradigm, the model is also tasked with predicting the hydrodynamic velocity field. The flow velocity in the x and y directions at the centre of cell i at timestep t, is denoted as u_i^t and v_i^t respectively. The velocity values are directly extracted from the simulation results and form the ground truth matrices $\mathcal{U}, \mathcal{V} \in \mathbb{R}^{N \times T}$.

By producing the entire time sequence for all cells as the prediction target, FloodTransformer operates in a non-autoregressive manner. This approach significantly accelerates the prediction process compared to traditional models that generate outputs step-by-step, making it exceptionally well-suited for real-time, high-resolution flood forecasting.

4.2. Variable-size cell representation

The core of our model is a multi-layer Transformer encoder. To enable the Transformer architecture to process the diverse input data, we transform the features associated with each of the N cells into a sequence of input representations, $\{\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_N\}$, where $\mathbf{Z}_i \in \mathbb{R}^D$ is a token representing cell i, and D is the hidden dimension of the model. The input representation $\mathbf{Z}_i$ is constructed by summing four distinct component embeddings: (1) a dynamic feature embedding, (2) a static feature embedding, (3) a positional embedding, and (4) a cell size embedding.

Dynamic Feature Embedding. This component encodes the time-varying inputs: the initial water state and the rainfall forecast. For each cell i, the initial water depth d_i^0 and inundation state h_i^0 are concatenated to form a 2-dimensional vector. Concurrently, the rainfall time series $\mathbf{r}_i \in \mathbb{R}^T$ is projected into a $D-2$ dimensional space using a dedicated Multi-Layer Perceptron ($\text{MLP}_{\mathcal{R}}$). These two resulting vectors are then concatenated to form the final dynamic feature embedding $\mathbf{z}_{\text{dyn},i} \in \mathbb{R}^D$, where

$$\mathbf{z}_{\text{dyn},i} = \text{concat}([d_i^0, h_i^0], \text{MLP}_{\mathcal{R}}(\mathbf{r}_i)). \tag{1}$$

Static Feature Embedding. The static geographical feature vector $\mathbf{s}_i \in \mathbb{R}^{D_s}$, which encapsulates the DEM, roughness, and infiltration information for cell i, is linearly projected into a D-dimensional space. This yields the static feature embedding $\mathbf{z}_{\text{sta},i} \in \mathbb{R}^D$.

Positional Embedding. To provide the model with explicit information about the spatial arrangement of the cells, we compute a positional embedding from the cell's centre coordinates. These coordinates are processed by a non-linear MLP, denoted as MLP_{pos}, to generate a learnable positional embedding $\mathbf{z}_{\text{pos},i} \in \mathbb{R}^D$.

Cell Size Embedding. Given that the cell schematisation may involve cells of varying sizes, we introduce a learnable embedding to encode this information. The study area contains C distinct cell area categories. We maintain a learnable embedding matrix $\mathbf{E}_{\text{size}} \in \mathbb{R}^{C \times D}$. For each cell i, we retrieve its corresponding cell size embedding $\mathbf{z}_{\text{size},i}$ from this matrix based on its area category.

Finally, the comprehensive input representation $\mathbf{Z}_i$ for each cell is obtained by the element-wise summation of these four component embeddings:

$$\mathbf{Z}_i = \mathbf{z}_{\text{dyn},i} + \mathbf{z}_{\text{sta},i} + \mathbf{z}_{\text{pos},i} + \mathbf{z}_{\text{size},i} \tag{2}$$

The resulting sequence of tokens, $[\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_N]$, which represents every single cell in the study area, serves as the input to the Transformer encoder. Within each layer of the encoder, a multi-head self-attention layer allows every cell (token) to weigh the influence of and integrate information from all other cells in the study area. This global receptive field is crucial for modelling the propagation of floodwaters and the non-local effects of rainfall and geographical features.

4.3. Multi-task training with regulations

The final sequence of cell representations from the Transformer encoder is fed into parallel prediction heads, each responsible for a distinct task: (1) regression of water depth, (2) regression of flow velocity, and (3) classification of the inundation state. The entire model, parameterised by θ, is trained end-to-end by minimising a composite loss function $\mathcal{L}$, which aggregates the individual losses from each of the three tasks.

Water Depth Regression Loss. The loss for the water depth regression, $\mathcal{L}_{\text{wd}}$, is composed of two distinct Mean Squared Error (MSE) (Hastie et al., 2009) terms to ensure both the accuracy of step-wise predictions and the stability of the cumulative forecast over the entire horizon.

First, to enforce precision at each individual timestep, we compute an MSE loss, $\mathcal{L}_{\text{step}}$, on the predicted variations of water depth (instead of the water depth directly) $\hat{\Delta d}_i^t$ as:

$$\mathcal{L}_{\text{step}} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (\hat{\Delta d}_i^t - \Delta d_i^t)^2, \tag{3}$$

where $\hat{\Delta d}_i^t$ and Δd_i^t are the predicted and ground truth water depth changes for cell i at timestep t, respectively.

Second, to mitigate the accumulation of errors over long sequences, we introduce an additional constraint, the final state loss $\mathcal{L}_{\text{final}}$, on the final water depth at the end of the forecast horizon, T, as:

$$\mathcal{L}_{\text{final}} = \frac{1}{NT} \sum_{i=1}^{N} (\hat{d_i^T} - d_i^T)^2, \tag{4}$$

where $\hat{d_i^T}$ is the predicted water depth at the end of the forecast horizon T and is calculated by accumulating the predicted variations from the initial state d_i^0:

$$\hat{d_i^T} = d_i^0 + \sum_{t=1}^{T} \hat{\Delta d}_i^t. \tag{5}$$

Thus, the Water Depth Regression loss is a sum of these two components:

$$\mathcal{L}_{\text{wd}} = \mathcal{L}_{\text{step}} + \lambda_{\text{final}} \cdot \mathcal{L}_{\text{final}}, \tag{6}$$

where λ_{final} is a hyperparameter that balances the importance of short-term and long-term accuracy.

Flow Velocity Regression Loss. To supervise the prediction of the flow velocity field, a regression loss, denoted as $\mathcal{L}_{\text{uv}}$, is formulated using MSE. The loss is the sum of the individual MSE losses for the x- (u) and y-direction (v) velocity components, defined as:

$$\mathcal{L}_{\text{uv}} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[(\hat{u_i^t} - u_i^t)^2 + (\hat{v_i^t} - v_i^t)^2 \right], \tag{7}$$

where $\hat{u}_i^t$ and $\hat{v}_i^t$ are the predicted velocity components for cell i at timestep t, while u_i^t and v_i^t are their corresponding ground truth values.

Inundation State Classification Loss. For the binary classification of inundation states, we employ a Binary Cross-Entropy (BCE) loss (Goodfellow et al., 2016), denoted as $\mathcal{L}_{\text{cls}}$. To address the significant class imbalance typically present between inundated (positive class) and non-inundated (negative class) cells, we introduce a weight, ω, for the positive class. The Inundation State Classification Loss is formulated as:

$$\mathcal{L}_{\text{cls}} = -\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \Big[\omega \cdot h_i^t \log(\sigma(\hat{y}_i^t)) + (1 - h_i^t) \log(1 - \sigma(\hat{y}_i^t)) \Big], \tag{8}$$

where $h_i^t \in \{0, 1\}$ is the ground truth inundation state, $\hat{y}_i^t$ is the raw logit output from the classification head for cell i at timestep t, and $\sigma(\cdot)$ is the sigmoid function.

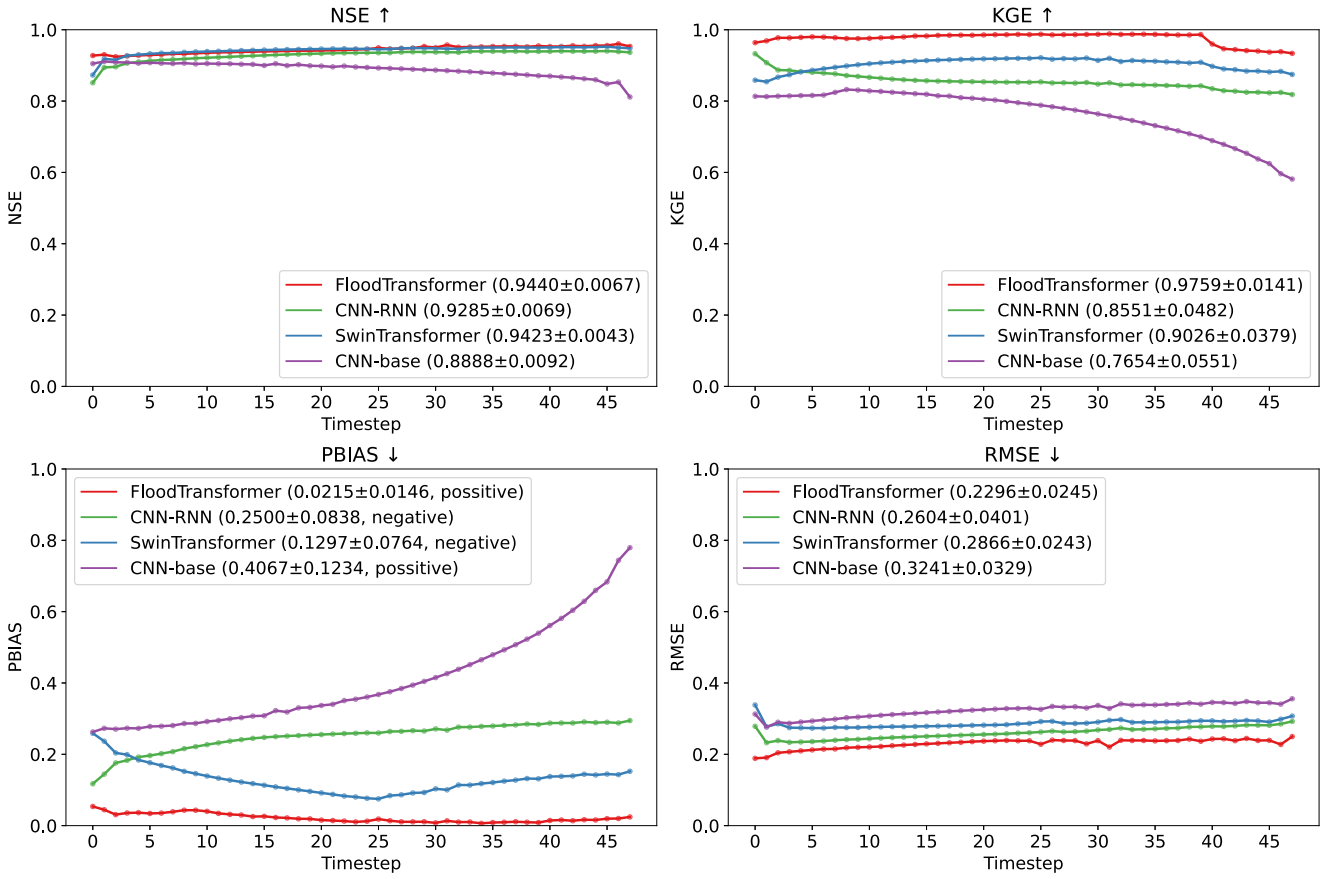


Fig. 7. Comparison results of the predicted water depth against high-resolution ground truth rasters for four models, *i.e.*, FloodTransformer (proposed), CNN-RNN (Alkaabi et al., 2025), SwinTransformer (Jamali and Mahdianpari, 2022), and CNN-base (Gao et al., 2024), using NSE, KGE, PBIAS, and RMSE metrics over multiple time steps at a 30-minute interval.

Total Loss. The final loss function is the weighted sum of the regression and classification losses:

$$\mathcal{L} = \mathcal{L}_{\text{wd}} + \lambda_{\text{uv}} \cdot \mathcal{L}_{\text{uv}} + \lambda_{\text{cls}} \cdot \mathcal{L}_{\text{cls}}, \tag{9}$$

where λ_{uv} and λ_{cls} are hyperparameters that control the contribution of the velocity regression and inundation classification tasks to the total loss. With Eq. (6), this can be further written as:

$$\mathcal{L} = \mathcal{L}_{\text{step}} + \lambda_{\text{final}} \cdot \mathcal{L}_{\text{final}} + \lambda_{\text{uv}} \cdot \mathcal{L}_{\text{uv}} + \lambda_{\text{cls}} \cdot \mathcal{L}_{\text{cls}}. \tag{10}$$

By requiring the model to predict multiple variables in a flood event, this training strategy constrains the model to learn representations that are more consistent with the underlying principles of fluid dynamics. The pseudocode in Algorithm 1 presents the FloodTransformer training procedure. The detailed model architecture hyperparameters and training hyperparameters are specified in Tables A.2 and A.3, respectively.

5. Results and evaluation

We evaluate the performance of the FloodTransformer in terms of both prediction accuracy and computational efficiency. For prediction, we assess the accuracy of the predicted water depth rasters (Fig. 7) and inundation state maps (Fig. 8) separately on both the simulation dataset and selected historical flood events in comparison with three other widely-used models of our implementation, *i.e.*, CNN-RNN (Alkaabi et al., 2025), SwinTransformer (Jamali and Mahdianpari, 2022), and CNN-base (Gao et al., 2024). All tests are conducted on a GPU server equipped with an NVIDIA L40 GPU.

All models are trained on the training set of the simulation dataset (detailed in Section 3.4), with hyperparameter tuning performed on the validation set. Tables A.2 and A.3 in Appendix A provide the model configuration details and training hyperparameters. Experiment results are assessed using 5-fold cross-validation on the training and validation sets, and evaluated on the held-out test set. Evaluation metrics (see Appendix B for detailed definitions) are reported as mean ± standard deviation in the following sections.

In addition to the simulation dataset, we further assess the FloodTransformer on three representative historical flood events in Wagga Wagga, using real historical rainfall and flood water level records. These case studies allow us to validate the model against real-world flood dynamics and test its generalisation capability beyond the simulated scenarios. The accuracy of the predicted inundation extents and water depth rasters is quantitatively compared with ground-truth measurements from gauging stations and high-resolution flood maps to comprehensively assess the performance of the proposed FloodTransformer.

5.1. Water depth prediction

In this section, we evaluate the performance of the water depth prediction using four widely used metrics in hydrology, including NSE (Gupta et al., 2009), KGE (Knoben et al., 2019), PBIAS (Gupta et al., 1999), and RMSE (Moriasi et al., 2015) (see Appendix B.1 for more details).

Fig. 7 shows the evaluation of predicted water depth against high-resolution ground truth rasters across 48 time steps (one day) at a 30-minute interval in comparison with the three other baseline models. As shown in the figure, across all evaluation metrics, FloodTransformer consistently ranks highest, demonstrating both superior accuracy and temporal stability. SwinTransformer shows competitive

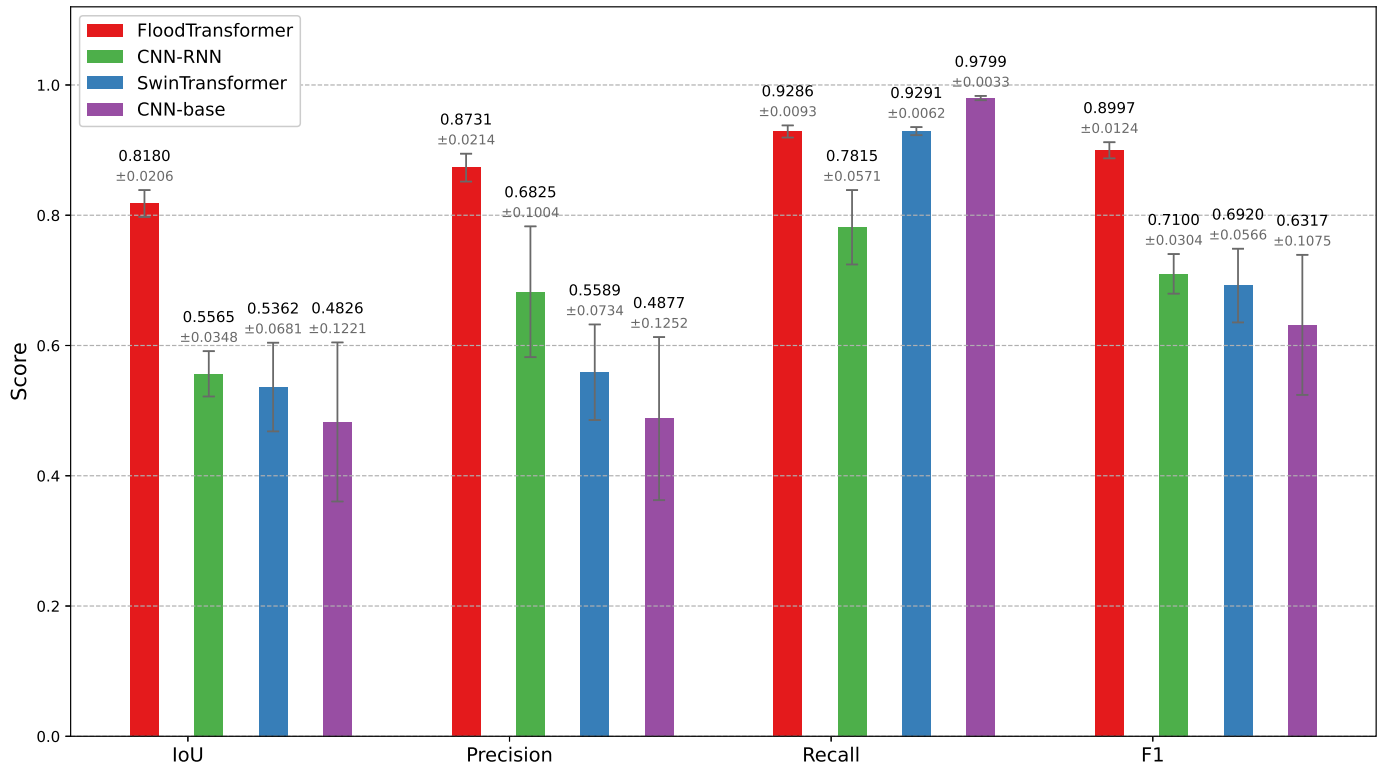


Fig. 8. Comparison of the inundation state predictions of FloodTransformer (proposed), CNN-RNN (Alkaabi et al., 2025), SwinTransformer (Jamali and Mahdianpari, 2022), and CNN-base (Gao et al., 2024), using IoU, Precision, Recall, and F1 score metrics over multiple time steps.

results in NSE and KGE but suffers from higher bias and error variability. CNN-RNN, which integrates physics-informed constraints, performs moderately well but exhibits greater fluctuations and systematic bias. CNN-base trails behind in all metrics, indicating limitations in capturing hydrodynamic patterns effectively.

In particular, FloodTransformer achieves a mean NSE of 0.9440, indicating excellent alignment with observed temporal dynamics. Its KGE score of 0.9759 further confirms high-quality predictions, capturing correlation, bias, and variability more comprehensively than other models. In terms of error metrics, FloodTransformer maintains a low RMSE of 0.2296, reflecting minimal deviations across timesteps. Most notably, its PBIAS is 0.0215 ± 0.0146, indicating a slight positive bias but significantly lower than the negative biases observed in CNN-RNN (0.2500 ± 0.0838) and SwinTransformer (0.1297 ± 0.0764), and the larger positive bias in CNN-base (0.4067 ± 0.1234).

Compared to the baselines, FloodTransformer's predictions are not only more accurate but also more consistent across the entire forecast horizon. Its consistent or slightly improved performance across all timesteps indicates that our method's single-run time sequence prediction effectively maintains temporal consistency, significantly reducing both the computational cost and cumulative error compared to traditional multi-step recursive forecasting approaches.

5.2. Inundation state prediction

For inundation state prediction (*i.e.*, distinguishing between flooded and non-flooded areas), we adopt widely used metrics from spatial classification tasks that are well-suited for evaluating inundation state, including IoU (Zeng et al., 2024), Precision, Recall, and the F1 score (Razali et al., 2020) (see Appendix B.2 for more details).

Fig. 8 compares inundation-state performance for the four comparative models: FloodTransformer (proposed), CNN-RNN, SwinTransformer, and CNN-base, using IoU, Precision, Recall, and F1. Overall, FloodTransformer consistently attains the highest scores on IoU, Precision, and F1, indicating superior predictive accuracy and temporal stability. Its low standard deviations across all metrics and high F1 indicate a stable, strong agreement between the predicted and observed inundation states. Both precision and recall remain consistently stable, with precision averaging slightly above 0.87 and recall maintaining values around 0.93. The F1 score remains robust and also has low deviation, confirming that the model effectively balances precision and recall over time for multiple flood events. These results underscore the model's reliability in flood inundation classification.

CNN-RNN ranks second, benefiting from hydrological priors and multi-task learning, though with moderate variability. SwinTransformer and CNN-base lag behind with lower scores and larger deviations. By contrast, CNN-base shows the weakest and least stable results, and SwinTransformer, despite its transformer backbone, struggles to maintain consistent precision and recall. These findings highlight FloodTransformer's effectiveness in capturing spatiotemporal flood dynamics and the limitations of conventional and vision-only baselines.

5.3. Historical flood events

We evaluate inundation-map predictions using three historical flood events with available ground-truth extent maps. Fig. 9 illustrates pixel-wise flood prediction performance for three distinct dates: 2022-10-24, 2022-10-08, and 2022-11-01. Each subfigure compares model predictions with ground truth data, colour-coded into four categories: true positives (TP, blue), true negatives (TN, light green), false positives (FP, red), and false negatives (FN, dark green).

We use the same inundation state evaluation metrics for historical flood events. Overall, they demonstrate excellent accuracy and consistency with all IoU values above 0.8 and F1 scores around 0.9. On 2022-10-24, the model achieved a precision of 0.8630, a recall of 0.9184, and an F1 score of 0.8898. On 2022-10-08, precision improved to 0.8743, with a recall of 0.9265 and an F1 score of 0.8997. The

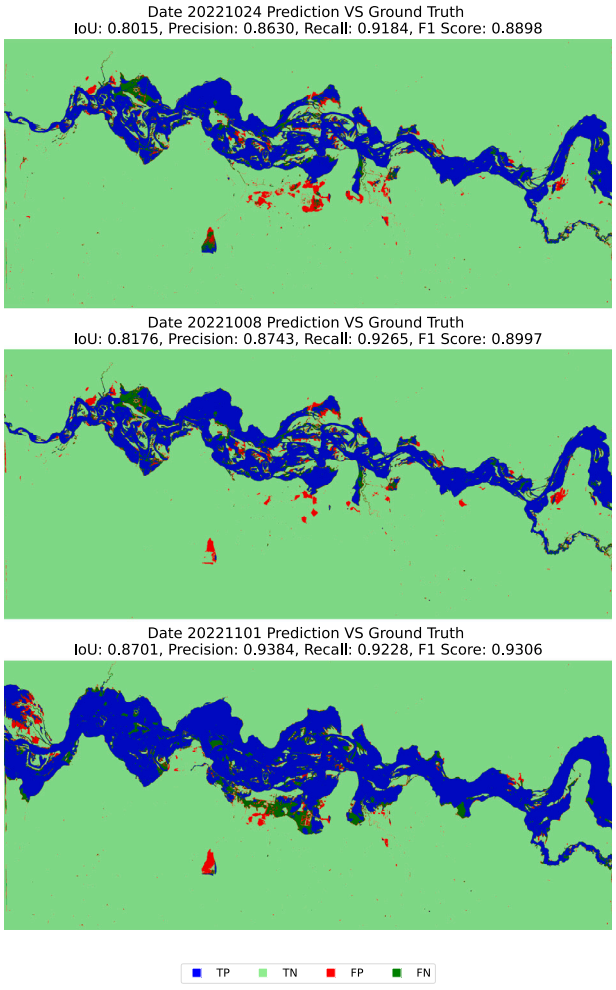


Fig. 9. Flood prediction results vs. ground truth for the three historical events. The events date and quantitative evaluation results are also reported.

highest performance was observed on 2022-11-01, with a precision of 0.9384, a recall of 0.9238, and an F1 score of 0.9306.

These results demonstrate strong predictive performance, particularly in accurately identifying flood-prone areas (high TP and low FN), though some overprediction remains visible in the form of scattered FP regions. The consistent spatial patterns across dates also suggest good model generalisation under varying flood conditions.

5.4. Comprehensive evaluation and efficiency analysis

In this section, we comprehensively evaluate the overall performance and computational efficiency of all models. A comparative analysis across multiple evaluation metrics is conducted to provide an integrated view of their predictive accuracy and robustness, along with computational statistics of the total number of parameters and TFLOPs to show the trade-off between model performance and efficiency.

We use a radar chart in Fig. 10 to provide a comprehensive comparison of four models across all evaluation metrics. FloodTransformer consistently dominates the chart, achieving near-maximum scores in all performance metrics, particularly excelling in Recall, F1, KGE, and NSE, which reflect its strong predictive accuracy and temporal consistency. CNN-RNN demonstrates competitive performance, benefiting from physics-informed constraints but exhibiting higher bias and error variability. SwinTransformer shows moderate results in NSE and KGE but underperforms in classification metrics like Precision and IoU. CNN-base ranks lowest overall, with weaker scores across both classification

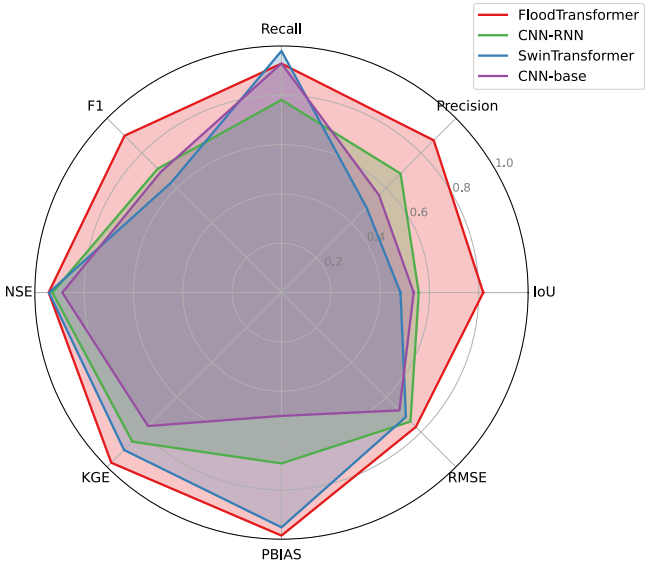


Fig. 10. Radar chart comparing four models' performance across all eight evaluation metrics.

and regression metrics, indicating limited effectiveness in capturing flood dynamics. This visual summary reinforces FloodTransformer's robustness and balanced performance across both spatial and temporal dimensions.

In addition to evaluating the prediction performance, we assess the computational efficiency of all models. Benchmark tests have been conducted on a GPU server equipped with an NVIDIA L40 GPU.

Table 3 presents the comparative analysis results of the computational efficiency of different models in terms of parameter size and theoretical floating-point operations per second (TFLOPs). Among the evaluated models, the FloodTransformer demonstrates a balanced trade-off between model complexity and computational demand, with 91.8803 million parameters and 92.9715 TFLOPs. This further demonstrates the effectiveness of our innovative designs in the model architecture, multi-task training, and optimisation strategies, as detailed in Section 4. In contrast, the SwinTransformer exhibits the highest computational cost, with 172.6023 million parameters and 39.6391 TFLOPs, reflecting its heavy architecture and substantial resource requirements yet comparatively lower performance. The CNN-RNN and CNN-base models are considerably lighter, with parameter counts of approximately 65–68 million and TFLOPs around 7.2700, demonstrating higher computational efficiency and suitability for low-resolution or resource-constrained applications. Overall, our proposed FloodTransformer achieves the best predictive performance while maintaining superior computational efficiency, making it ideal for high-resolution and high-accuracy flood forecasting tasks with optimal utilisation of computational resources. This represents a dramatic improvement over conventional hydrodynamic models, which often require hours to complete comparable simulations.

6. Discussion

In this section, we discuss the limitations of our proposed FloodTransformer, despite its excellent performance and high efficiency, and outline strategies and future works to further improve its accuracy, robustness, and practical applicability in real-world flood forecasting scenarios.

While our approach demonstrates strong predictive performance across river channels, surrounding areas, and broad ungauged regions, it does not incorporate specialised modelling or treatment for urban flood dynamics. Similar to many existing studies (Xu and Gao, 2024;

Table 3
Model comparison on parameter size, total inferencing time for one day lead time, and theoretical floating-point operations per second (TFLOPs). The TFLOPs indicate the computational capacity of a model, reflecting the number of floating-point calculations it can perform per second.

Model	Total Parameters (Million)	Total Inferencing time (s)	TFLOPs
FloodTransformer	91.8803	2.9766	92.9715
CNN-RNN	67.7743	13.9910	7.2773
SwinTransformer	172.6023	5.6973	39.6391
CNN-base	64.6265	13.8986	7.2664

Zhong et al., 2024), our current hydrodynamic modelling does not explicitly represent urban drainage and sewer networks.

This limitation arises from the inherent complexity and sensitivity of these systems, as well as the high uncertainty associated with their operational states, such as potential blockages and fluctuating usage due to daily human activities. These factors significantly increase the modelling complexity and make accurate simulation challenging. However, this limitation related to the physical models is not believed to significantly affect their high flow modelling outcomes where the AI-based model focuses on. To address this, our following work will enhance the simulation stage by integrating explicit urban drainage components within the hydrodynamic model and coupling them with AI-based learning to better capture fine scale urban flood behaviour. This integration will further improve the modelling accuracy in densely built environments and enables more reliable urban flood forecasting.

In addition, although this study focuses on a specific region, the proposed FloodTransformer framework shows strong potential for wider application across other catchments and flood-prone areas. With its data-driven and physics-informed architecture, the system is well-suited for adaptation to different hydrological settings. However, given the variability in terrain, land use, infrastructure, and rainfall-runoff responses across different regions, further testing and training are necessary to fully assess the robustness and performance of the approach. To enhance transferability, in future research we will employ domain adaptation and transfer learning techniques for efficient retraining across new catchments, supported by automated batch simulations to reduce the workload compared to developing entirely new hydrodynamic and AI models. Additionally, we will incorporate additional data modalities, such as remote sensing and real-time sensor observations, to improve data quality and ensure operational reliability. Moreover, for practical deployment, a promising way is to integrate the framework with regional forecasting systems and emergency management platforms to address real-world challenges, including computational constraints and data accessibility.

While this study adopts a multi-task training framework with physics-informed regulations, we do not explicitly enforce mass conservation or the shallow-water momentum equations. In future work, we will incorporate more physics penalties that strengthen physical consistency: (i) a domain-wise volume-conservation loss that matches the integrated volume change to the net rainfall–infiltration input and boundary fluxes over the forecast horizon; and (ii) boundary-flow consistency penalties aligning predicted fluxes along 2D boundary lines with upstream/downstream conditions. These terms can be adaptively weighted and provide a practical path towards stronger physics-informed learning without sacrificing runtime.

7. Conclusion

In this study, we developed an effective AI-hydrodynamic hybrid flood forecasting framework that integrates physics-based hydrodynamic simulations with an efficient and novel deep learning model, FloodTransformer. Training on large-volume, diverse simulations generated from a calibrated 3Di hydrodynamic model, our approach effectively bridges the gap between time-consuming hydrodynamic calculations and real-time prediction requirements.

The FloodTransformer provides multiple innovative designs, including multimodal data processing, variable-size cell embeddings, and tokenised time-sequence encodings, to efficiently capture complex spatiotemporal dynamics across large-scale and high-resolution flood scenarios. The integration of physics-informed multi-task optimisation ensures physical consistency in forecasting, enhancing the model's reliability for real-world applications. Furthermore, this framework enables real-time sequential predictions on water depths and flood inundation maps, while significantly reducing computational overhead and aggregated errors compared to traditional iterative approaches. Extensive experiments and evaluations on both simulated and historical flood events confirm that our framework achieves high accuracy and efficiency across a wide range of flood scenarios.

In practical terms, the real-world implementation will focus on operational integration with government early warning systems and hydrological data infrastructures such as those managed by the BoM and local emergency services. The engaging stakeholders including water authorities, urban planners, and community responders will be essential to ensure seamless deployment and user accessibility. In future research, we will extend the framework to larger-scale and cross-catchment applications, integrating additional data modalities such as real-time sensor observations and employing transfer learning strategies to enhance adaptability in new regions.

CRediT authorship contribution statement

Zhanzhong Gu: Writing – review & editing, Writing – original draft, Visualisation, Software, Methodology, Formal analysis, Data curation, Conceptualisation. **Jiachen Kang:** Writing – review & editing, Writing – original draft, Visualisation, Software, Methodology, Formal analysis, Conceptualisation. **Wenzheng Jin:** Validation, Software, Methodology, Data curation, Conceptualisation. **Feifei Tong:** Writing – review & editing, Validation, Methodology. **Y. Jay Guo:** Writing – review & editing, Validation, Supervision, Resources, Conceptualisation. **Wenjing Jia:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Methodology, Conceptualisation.

Ethics

This study is undertaken strictly in compliance with the Australia National Statement on Ethical Conduct and UTS Research Policy.

Software and data availability

- Name of software: 3Di.
  - Developer: Nelen & Schuurmans.
  - Contact: info@3diwatermanagement.com.
  - Date first available: March 23, 2021.
  - Software required: QGIS.
  - Programming language: Python, SQL, JavaScript.
  - Source code: https://github.com/nens.
  - Documentation: Detailed documentation for installation, testing, and deployment can be found at https://docs.3di.live/.
- Name of software: FloodTransformer.
  - Developer: Jiachen Kang, Zhanzhong Gu, Wenzheng Jin
  - Contact: Jiachen.Kang-1@uts.edu.au, Zhanzhong.gu@uts.edu.au, Wenzheng.Jin@uts.edu.au.
  - Date first available: August 14, 2025.

Table A.1
Notation summary.

Symbol	Meaning
$\mathcal{A}$	Study area (spatial domain).
N	Number of computational cells.
i	Index of computational cell, $i \in \{1, \dots, N\}$.
t	Timestep index.
T	Forecast horizon (number of timesteps).
F_θ	FloodTransformer mapping from inputs to outputs, parameterised by θ.
θ	Model parameters.
$\mathcal{X}$	Input predictive variables to the model.
$\mathcal{Y}_{\text{sim}}$	Ground-truth flood states from hydrodynamic simulations.
$\hat{\mathcal{Y}}$	Model-predicted flood states.
$\mathcal{S}$	Set of static feature vectors.
$\mathbf{s}_i \in \mathbb{R}^{D_s}$	Static feature vector for cell i; D_s is the static feature dimension.
d_i^t	Water depth at cell i and timestep t.
$\mathbf{d}^0$	Initial water depth vector across all cells.
h_i^t	Inundation state (binary) at cell i and timestep t.
$\mathbf{h}^0$	Initial inundation state vector.
$\mathcal{R}$	Rainfall time-series matrix over all cells.
$\mathbf{r}_i$	Rainfall time series for cell i over the horizon.
r_i^t	Rainfall at cell i and timestep t.
ΔD, $\Delta\hat{D}$	Ground truth and predicted matrix of water depth changes over cells and timesteps.
Δd_i^t	Water depth change at cell i and timestep t.
$\mathcal{H}$, $\hat{\mathcal{H}}$	Ground truth and predicted inundation state matrix over cells and timesteps.
$\mathcal{U}, \mathcal{V}$, $\hat{\mathcal{U}}, \hat{\mathcal{V}}$	Ground truth and predicted flow velocity matrices in x and y directions.
u_i^t, v_i^t	Flow velocity components at cell i, timestep t.
D	Transformer hidden dimension.
$\mathbf{Z}_i$	Token representation for cell i.
$\mathbf{z}_{\text{dyn},i}$, $\mathbf{z}_{\text{sta},i}$, $\mathbf{z}_{\text{pos},i}$, $\mathbf{z}_{\text{size},i}$	Dynamic feature, static feature, positional and cell-size embedding for cell i.
C	Number of cell-size categories.
L	Number of Transformer encoder layers.
H	Number of attention heads.
$\mathcal{L}_{\text{step}}$	Step-wise depth-change loss.
$\mathcal{L}_{\text{final}}$	Final-depth loss at the end of the horizon T.
$\mathcal{L}_{\text{wd}}$	Water-depth regression loss.
$\mathcal{L}_{\text{uv}}$	Velocity regression loss.
$\mathcal{L}_{\text{cls}}$	Inundation classification loss.
ω	Positive-class weight for binary cross-entropy.
$\mathcal{L}$	Total training loss.
$\lambda_{\text{final}}, \lambda_{\text{uv}}, \lambda_{\text{cls}}$	Loss weights for final-depth, velocity, and classification terms.

- Software required: Python, Linux.
- Program language: Python.
- Source code at: https://github.com/jiachenkang/FloodTransformer.
- Documentation: Detailed documentation for the installation, testing, and deployment can be found at https://github.com/jiachenkang/FloodTransformer/blob/main/README.md.
- Data: The test data can be downloaded from https://docs.google.com/forms/d/e/1FAIpQLScBgqVwQiAjWtKkI2HvgQndPHzn8yZ4ps9LzTZNSKLBQrqKag/viewformwith password 'ft1e'.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the New South Wales (NSW) State Emergency Service (SES) via the NSW Digital Restart Fund. We gratefully acknowledge the strong support provided by NSW SES, particularly Stephen McRobert and David Monk. Additionally, we acknowledge and appreciate the licenses and consultancy services provided by Nelen & Schuurmans that enabled our access to the 3Di API platform.

Appendix A. FloodTransformer: Additional implementation and training details

In support of Section 4, this section includes the training pseudo-code, key model configuration, and the training hyper-parameters used in our experiments. All notations are preserved and summarised in Table A.1, so references in the main text remain valid.

A.1. Training procedure pseudo-code

The full training procedure for FloodTransformer is presented in Algorithm 1 for clarity and implementation reference.

A.2. Key configuration details

Table A.2 summarises core architectural settings that define the FloodTransformer capacity and token dimensions used throughout Section 4.

A.3. Training hyperparameters

Table A.3 lists optimiser, schedule, precision, and loss weights to facilitate reproduction of results.

Input: Static features $S = \{\mathbf{s}_1, \ldots, \mathbf{s}_N\}$; Initial water depth $\mathbf{d}^0 = [d_1^0, \ldots, d_N^0]$; Initial inundation $\mathbf{h}^0 = [h_1^0, \ldots, h_N^0]$ where $h_i^0 = \mathbb{I}(d_i^0 > 0)$; Rainfall $\mathcal{R} = \{\mathbf{r}_1, \ldots, \mathbf{r}_N\}$ where $\mathbf{r}_i = [r_i^1, \ldots, r_i^T] \in \mathbb{R}^T$

Output: Water depth change $\hat{\Delta D} \in \mathbb{R}^{N\times T}$; Inundation state $\hat{\mathcal{H}} \in \mathbb{R}^{N\times T}$; Flow velocity $\hat{\mathcal{U}}, \hat{\mathcal{V}} \in \mathbb{R}^{N\times T}$

Learn model $\mathcal{F}_\theta$ to predict flood evolution for N cells over T timesteps;

// Initialisation (training phase)

Initialise Transformer encoder with L layers, H heads, dimension D;

Initialise heads: Head_{wd}, Head_{cls}, Head_{u}, Head_{v};

Initialise size embeddings with C categories: $\mathbf{E}_{\text{size}} \in \mathbb{R}^{C\times D}$;

// Training

for *each epoch* **do**

  for *each batch* **do**

    // Tokenized time-sequence encoding

    for $i = 1$ *to* N **do**

      $\mathbf{z}_{\text{rain},i} = \text{MLP}_{\mathcal{R}}(\mathbf{r}_i)$;

      $\mathbf{z}_{\text{dyn},i} = \text{concat}([d_i^0, h_i^0], \mathbf{z}_{\text{rain},i})$;

      $\mathbf{z}_{\text{sta},i} = \text{Linear}(\mathbf{s}_i)$;

      $\mathbf{z}_{\text{pos},i} = \text{MLP}_{\text{pos}}(\text{coords}_i)$;

      $\mathbf{z}_{\text{size},i} = \mathbf{E}_{\text{size}}[\text{category}_i]$;

      $\mathbf{Z}_i = \mathbf{z}_{\text{dyn},i} + \mathbf{z}_{\text{sta},i} + \mathbf{z}_{\text{pos},i} + \mathbf{z}_{\text{size},i}$;

    end

    $\mathbf{X} = [\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_N] \in \mathbb{R}^{N\times D}$;

    $\mathbf{X} = \text{LayerNorm}(\text{Transformer}(\mathbf{X}))$;

    // Non-autoregressive multi-task prediction

    $\hat{\Delta D} = \text{Head}_{\text{wd}}(\mathbf{X}) \in \mathbb{R}^{N\times T}$;

    $\hat{\mathcal{H}} = \text{Head}_{\text{cls}}(\mathbf{X}) \in \mathbb{R}^{N\times T}$;

    $\hat{\mathcal{U}} = \text{Head}_{\text{u}}(\mathbf{X}) \in \mathbb{R}^{N\times T}$;

    $\hat{\mathcal{V}} = \text{Head}_{\text{v}}(\mathbf{X}) \in \mathbb{R}^{N\times T}$;

    // Physics-informed multi-task loss

    $\mathcal{L}_{\text{step}} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\hat{\Delta d}_i^t - \Delta d_i^t)^2$;

    $\hat{d}_i^T = d_i^0 + \sum_{t=1}^{T}\hat{\Delta d}_i^t$ for $i = 1, \ldots, N$;

    $\mathcal{L}_{\text{final}} = \frac{1}{NT}\sum_{i=1}^{N}(\hat{d}_i^T - d_i^T)^2$;

    $\mathcal{L}_{\text{wd}} = \mathcal{L}_{\text{step}} + \lambda_{\text{final}} \cdot \mathcal{L}_{\text{final}}$;

    $\mathcal{L}_{\text{uv}} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[(\hat{u}_i^t - u_i^t)^2 + (\hat{v}_i^t - v_i^t)^2\right]$;

    $\omega = \frac{\#\text{ non-inundated}}{\#\text{ inundated}}$;

    $\mathcal{L}_{\text{cls}} = \text{WeightedBCE}(\hat{\mathcal{H}}, \mathcal{H}, \omega)$;

    $\mathcal{L} = \mathcal{L}_{\text{wd}} + \lambda_{\text{uv}} \cdot \mathcal{L}_{\text{uv}} + \lambda_{\text{cls}} \cdot \mathcal{L}_{\text{cls}}$;

    $\theta \leftarrow \text{Optimiser}(\theta, \nabla_\theta \mathcal{L})$;

  end

end

Algorithm 1: The pseudo-code of the FloodTransformer training procedure.

Table A.2
Key configuration details of the proposed FloodTransformer.

Parameter	Symbol	Value
Number of cells	N	47,791
Hidden dimension	D	768
Number of Transformer layers	L	12
Number of attention heads	H	12
Prediction timesteps	T	48
DEM embedding dimension	D_s	1280
Cell size categories	C	5
MLP expansion ratio	–	4
Rain embedding hidden dim	–	512
Position embedding hidden dim	–	512

Table A.3
Hyperparameters used for training the FloodTransformer.

Hyperparameter	Value
Optimisation	
Number of epochs	400
Initial learning rate	1×10^{-4}
Last learning rate	5×10^{-6}
LR scheduler	Cosine Annealing
Weight decay	1×10^{-3}
Optimiser	AdamW ($\beta_1 = 0.9$, $\beta_2 = 0.98$)
Data type	bfloat16
Loss Weights	
λ_{final}	1.0
λ_{uv}	0.5
λ_{cls}	1.0
BCE positive class weight	$11/89 \approx 0.124$

Appendix B. Evaluation metrics

In this section, we provide detailed definitions for the evaluation metrics adopted in this work, including the widely used hydrological metrics for water depth prediction and spatial classification metrics for inundation state prediction. In this section, we provide a detailed explanation of the evaluation metrics used in this study, including both the widely used hydrological metrics for water depth prediction and the spatial classification metrics for inundation state prediction.

B.1. Metrics for water depth prediction

To ensure a thorough and objective evaluation of the flood forecasting model's performance, four well-established statistical metrics commonly used in hydrological modelling were applied. These include the Nash–Sutcliffe Efficiency (NSE) (Gupta et al., 2009), which measures how well the predicted time series matches the observed data; the Kling–Gupta Efficiency (KGE) (Knoben et al., 2019), which provides a balanced assessment of correlation, bias, and variability errors; the Percent Bias (PBIAS) (Gupta et al., 1999), which quantifies the average tendency of simulated values to overestimate or underestimate the observed values; and the Root Mean Square Error (RMSE) (Moriasi et al., 2015), which emphasises larger deviations between simulated and observed values.

Nash–Sutcliffe Efficiency (NSE) measures how well predicted values replicate observed values, defined as:

$$\text{NSE} = 1 - \frac{\sum_{i=1}^{n}(s_i - o_i)^2}{\sum_{i=1}^{n}(o_i - \bar{o})^2}, \tag{11}$$

where s_i and o_i are the simulated and observed values, and $\bar{o}$ is the mean of observed values. NSE ranges from $-\infty$ to 1.0, where 1.0 indicates a perfect match, 0 indicates that the model is as accurate as the mean of the observations, and values less than 0 imply that the model performs worse than simply using the average of observed data.

Kling–Gupta Efficiency (KGE) provides a balanced assessment by incorporating correlation, bias, and variability, particularly useful for capturing overall agreement between predicted and observed time series, calculated as:

$$\text{KGE} = 1 - \sqrt{(r-1)^2 + (\beta-1)^2 + (\gamma-1)^2}, \tag{12}$$

where r is the Pearson correlation coefficient, $\beta = \frac{\bar{s}}{\bar{o}}$ is the bias ratio, and $\gamma = \frac{CV_s}{CV_o}$ is the variability ratio, with CV denoting the coefficient of variation. KGE also ranges from $-\infty$ to 1.0, with 1.0 being optimal. Values closer to 1 indicate better performance across all three components.

Percent Bias (PBIAS) quantifies the average tendency of simulated data to be larger or smaller than their corresponding observed values,

expressed as a percentage, indicating whether a model systematically overestimates or underestimates the observations:

$$\text{PBIAS} = 100 \times \frac{\sum_{i=1}^{n}(o_i - s_i)}{\sum_{i=1}^{n} o_i}. \tag{13}$$

A positive PBIAS value indicates that the model tends to *underestimate* the observations, whereas a negative value implies *overestimation*. A PBIAS value of zero represents a perfect model with no bias. According to Moriasi et al. (2007), PBIAS values within ±10% are considered *very good*, reflecting minimal systematic deviation between simulated and observed data.

Root Mean Square Error (RMSE) measures the average magnitude of model prediction errors, providing an overall measure of model accuracy by comparing simulated and observed values:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(s_i - o_i)^2}. \tag{14}$$

RMSE is non-negative, with 0 being ideal. However, because it squares the errors, RMSE is particularly sensitive to large discrepancies, making it effective for detecting extreme deviations. Lower RMSE values imply better model performance. These metrics provide robust assessments of predictive water depth against the ground truth, capturing both overall agreement and detailed error characteristics.

B.2. Metrics for inundation state prediction

For the evaluation of inundation state prediction, specifically the classification of flooded versus non-flooded areas, this study employs a set of widely recognised performance metrics commonly used in spatial classification and segmentation tasks. These include the Intersection over Union (IoU) (Zeng et al., 2024), which measures the overlap between predicted and observed flood extents; Precision and Recall, which respectively quantify the accuracy and completeness of flood detection; and the F1 score (Razali et al., 2020), which provides a harmonic balance between Precision and Recall.

IoU quantifies the spatial overlap between the predicted inundation state and the ground truth, defined as: $\text{IoU} = \frac{\text{TP}}{\text{TP+FP+FN}}$, where TP is the number of true positives (correctly predicted flooded) cells, FP is the number of false positives, and FN is the number of false negatives. IoU ranges from 0 to 1, where 1 indicates perfect spatial agreement.

Precision measures the proportion of correctly identified flooded cells among all predicted flooded cells, and is defined as:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{15}$$

A high precision means fewer false alarms in flood prediction, with a value range of $[0, 1]$.

Recall assesses the model's ability to detect all actual flooded cells:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{16}$$

A high recall indicates fewer missed flooded areas. Like precision, it ranges from 0 (no flooded cells detected) to 1 (all flooded cells detected).

F1 Score is the harmonic mean of Precision and Recall, providing a balanced measure that accounts for both false positives and false negatives:

$$\text{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{17}$$

F1 score also ranges from 0 to 1 and is valuable in class-imbalanced scenarios, especially in flood events where the inundation ratio changes dramatically throughout the events.

Table A.4
Ablation study on the contribution of individual loss components.

$\mathcal{L}_{\text{cls}}$	$\mathcal{L}_{\text{final}}$	$\mathcal{L}_{\text{uv}}$	RMSE	f1
×	×	×	0.3847	0.8580
✓	×	×	0.2405	0.9027
✓	✓	×	0.2554	0.8977
✓	✓	✓	**0.2296**	**0.8997**

Table A.5
Sensitivity analysis of the classification loss weight λ_{cls}. $\lambda_{\text{cls}} = 1$ achieves the optimal accuracy in water depth regression.

λ_{cls}	RMSE	f1
0.5	0.3230	0.8759
1	**0.2296**	**0.8997**
2	0.2723	0.9017
4	0.2813	0.8858

Appendix C. Ablation studies results

To systematically evaluate the contribution of each loss component, we conduct ablation studies on the three primary loss terms: $\mathcal{L}_{\text{cls}}$, $\mathcal{L}_{\text{final}}$, and $\mathcal{L}_{\text{uv}}$. In this section, we include these ablation studies results detailing component-wise contributions and sensitivity analyses. These results support the claims on model performance reported in the main text.

C.1. Contribution of individual loss components

Table A.4 presents the results examining the contribution of each loss component. The introduction of $\mathcal{L}_{\text{cls}}$ substantially improves both RMSE and F1 scores. This enhancement can be attributed to the multi-task learning framework, where the auxiliary classification task enforces a clear decision boundary at zero water depth, which helps regularise the regression predictions and reduces ambiguity in shallow water scenarios where depth estimation is inherently more challenging.

Interestingly, the addition of $\mathcal{L}_{\text{final}}$ results in a slight increase in RMSE. The conflict between the final-step loss and the step-wise loss forces the model to redistribute prediction errors across the sequence, leading to a modest degradation in average step-wise accuracy. In practice, the final-step water depth determines the ultimate extent of flooding, making it more important than intermediate states for decision-making purposes.

The incorporation of the flow velocity loss $\mathcal{L}_{\text{uv}}$ yields further improvement, with RMSE decreasing to 0.2296. This enhancement is grounded in the physical coupling between water depth and flow velocity in flood dynamics. By requiring the model to simultaneously predict both quantities, we implicitly enforce consistency with these underlying physical principles, thereby improving the overall prediction accuracy.

C.2. Sensitivity analysis for λ_{cls}

Table A.5 reveals that $\lambda_{\text{cls}} = 1$ achieves the optimal prediction accuracy. While insufficient weight ($\lambda_{\text{cls}} = 0.5$) fails to provide adequate regularisation for the depth regression, increasing λ_{cls} to 2 or 4 leads to worse RMSE, despite marginal improvements for F1 score. This degradation at higher λ_{cls} values can be attributed to the competing objectives in multi-task learning, where the binary supervision provides limited gradient information for the regression task.

Table A.6
Sensitivity analysis of the velocity loss weight λ_{uv}. λ_{uv} = 0.5 provides the optimal model performance.

λ_{uv}	RMSE	f1
0.25	0.2415	0.8985
0.5	**0.2296**	**0.8997**
1	0.2316	0.9011

C.3. Sensitivity analysis for λ_{uv}

Velocity-loss weighting primarily provides a modest regularising effect; λ_{uv} = 0.5 gives the best balance under our setup, as shown in Table A.6.

The results indicate that the $\mathcal{L}_{uv}$ has a relatively modest effect on overall performance. Increasing λ_{uv} from 0.25 to 1 yields a marginal improvement in RMSE. The limited impact of λ_{uv} can be attributed to a softer and indirect constraint provided by $\mathcal{L}_{uv}$ compared to $\mathcal{L}_{cls}$.

Data availability

I have shared the link to my data/code at the Attach File step.

References

3Di, 2025. 3Di water management. https://3diwatermanagement.com/. (Accessed 12 June 2025).

ABARES, 2025. Australian bureau of agricultural and resource economics and sciences (ABARES). https://www.agriculture.gov.au/abares. (Accessed 12 June 2025).

Adnan, M.S.G., Siam, Z.S., Kabir, I., Kabir, Z., Ahmed, M.R., Hassan, Q.K., Rahman, R.M., Dewan, A., 2023. A novel framework for addressing uncertainties in machine learning-based geospatial approaches for flood prediction. J. Environ. Manag. 326, 116813.

Al-Rawas, G., Nikoo, M.R., Al-Wardy, M., Etri, T., 2024. A critical review of emerging technologies for flash flood prediction: Examining artificial intelligence, machine learning, internet of things, cloud computing, and robotics techniques. Water 16, 2069.

Alkaabi, K., Sarfraz, U., Al Darmaki, S., 2025. A deep learning framework for flash-flood-runoff prediction: Integrating CNN-RNN with neural ordinary differential equations (ODEs). Water 17, 1283.

Anik, M.S.B.M., An, C., Li, S.S., 2025. Evolution from the physical process-based approaches to machine learning approaches to predicting urban floods: A literature review. Environ. Syst. Res. 14, 15.

Antwi-Agyakwa, K.T., Afenyo, M.K., Angnuureng, D.B., 2023. Know to predict, forecast to warn: A review of flood risk prediction tools. Water 15, 427.

Ball, J.E., Babister, M.K., Nathan, R., Weinmann, P.E., Weeks, W., Retallick, M., Testoni, I., 2016. Australian rainfall and runoff—A guide to flood estimation.

Biswakalyani, C., Samantaray, S., Satpathy, D.P., 2024. Assessment of spatio-temporal characteristics of groundwater quality at Khordha District, India. Discov. Civ. Eng. 1, 28.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Byaruhanga, N., Kibirige, D., Gokool, S., Mkhonta, G., 2024. Evolution of flood prediction and forecasting models for flood early warning systems: A scoping review. Water 16, 1763.

Campbell, F.B., Bauder, H.A., 1940. A rating-curve method for determining silt-discharge of streams. Eos Trans. Am. Geophys. Union 21, 603–607.

Castangia, M., Grajales, L.M.M., Aliberti, A., Rossi, C., Macii, A., Macii, E., Patti, E., 2023. Transformer neural networks for interpretable flood forecasting. Environ. Model. Softw. 160, 105581.

Chand, R., Deo, R.C., Ghimire, S., Nguyen-Huy, T., Ali, M., 2025. Hybrid CNN–GRU model for hourly flood forecasting index: Case studies from the Fiji islands. Stoch. Environ. Res. Risk Assess. 1–27.

Chang, Y.-S., Ho, H.-C., Huang, L.-Y., 2023. Evaluation of low impact development for resilient capacity in urban area with flood resilience index. J. Flood Risk Manag. 16, e12877.

Chen, J., Li, Y., Zhang, C., Tian, Y., Guo, Z., 2023. Urban flooding prediction method based on the combination of LSTM neural network and numerical model. Int. J. Environ. Res. Public Health 20, 1043.

Chowdhury, M.E., Islam, A.K.M.S., Lemans, M., Hegnauer, M., Sajib, A.R., Pieu, N.M., Das, M.K., Shadia, N., Haque, A., Roy, B., et al., 2023. An efficient flash flood forecasting system for the un-gaged Meghna basin using open source platform Delft-FEWS. Environ. Model. Softw. 161, 105614.

Copernicus Emergency Management Service, 2024a. European flood awareness system (EFAS). https://www.copernicus.eu/en/european-flood-awareness-system. (Accessed 12 June 2025).

Copernicus Emergency Management Service, 2024b. Global flood awareness system (glofas). https://global-flood.emergency.copernicus.eu/. (Accessed 12 June 2025).

Cox, N.J., 2018. Speaking stata: From rounding to binning. Stata J. 18, 741–754.

Dahm, R., Hsu, C.-T., Lien, H.-C., Chang, C.-H., Prinsen, G., 2014. Next generation flood modelling using 3Di: A case study in Taiwan. In: DSD International Conference.

Dang, T.Q., Tran, B.H., Le, Q.N., Tanim, A.H., Bui, V.H., Mai, S.T., Thanh, P.N., Anh, D.T., 2025. Integrating intelligent hydro-informatics into an effective early warning system for risk-informed urban flood management. Environ. Model. Softw. 183, 106246.

Do Lago, C.A.F., Giacomoni, M.H., Bentivoglio, R., Taormina, R., Junior, M.N.G., Mendiondo, E.M., 2023. Generalizing rapid flood predictions to unseen urban catchments with conditional generative adversarial networks. J. Hydrol. 618, 129276.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N., 2021. An image is worth 16 × 16 words: Transformers for image recognition at scale. Int. Conf. Learn. Represent..

Elsadek, W.M., Wahba, M., Al-Arifi, N., Kanae, S., El-Rawy, M., 2024. Scrutinizing the performance of GIS-based analytical hierarchical process approach and frequency ratio model in flood prediction–case study of kakegawa, Japan. Ain Shams Eng. J. 15, 102453.

Feng, D., Tan, Z., Lin, Z., Xu, D., Yu, C.-W., He, Q., 2025. A comparative study of physics-informed and data-driven neural networks for compound flood simulation at river-ocean interfaces: A case study of hurricane irene. J. Geophys. Res.: Mach. Learn. Comput. 2, e2025JH000758.

Gao, W., Liao, Y., Chen, Y., Lai, C., He, S., Wang, Z., 2024. Enhancing transparency in data-driven urban pluvial flood prediction using an explainable CNN model. J. Hydrol. 645, 132228.

Geoscience Australia, 2015. 5 metre digital elevation model (DEM) of Australia derived from LiDAR. https://pid.geoscience.gov.au/dataset/ga/89644. (Accessed 12 June 2025).

Gharakhanlou, N.M., Perez, L., 2023. Flood susceptible prediction through the use of geospatial variables and machine learning methods. J. Hydrol. 617, 129121.

Goodfellow, I., Bengio, Y., Courville, A., Bengio, Y., 2016. Deep Learning, vol. 1. MIT press Cambridge.

Guo, Z., Moosavi, V., Leitão, J.P., 2022. Data-driven rapid flood prediction mapping with catchment generalizability. J. Hydrol. 609, 127726.

Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. J. Hydrol. 377, 80–91.

Gupta, H.V., Sorooshian, S., Yapo, P.O., 1999. Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration. J. Hydrol. Eng. 4, 135–143.

Hakim, D.K., Gernowo, R., Nirwansyah, A.W., 2024. Flood prediction with time series data mining: Systematic review. Nat. Hazards Res. 4, 194–220.

Hastie, T., Tibshirani, R., Friedman, J.H., Friedman, J.H., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, vol. 2. Springer.

Hayder, I.M., Al-Amiedy, T.A., Ghaban, W., Saeed, F., Nasser, M., Al-Ali, G.A., Younis, H.A., 2023. An intelligent early flood forecasting and prediction leveraging machine and deep learning algorithms with advanced alert system. Processes 11, 481.

Hearst, M.A., Dumais, S.T., Osuna, E., Platt, J., Scholkopf, B., 1998. Support vector machines. IEEE Intell. Syst. Appl. 13, 18–28.

Hu, F., Yang, Q., Yang, J., Shao, J., Wang, G., 2025. An adaptive rainfall-runoff model for daily runoff prediction under the changing environment: Stream-LSTM. Environ. Model. Softw. 106524.

Huynh, N.N.T., Garambois, P.-A., Renard, B., Colleoni, F., Monnier, J., Roux, H., 2025. A distributed hybrid physics–AI framework for learning corrections of internal hydrological fluxes and enhancing high-resolution regionalized flood modeling. Hydrol. Earth Syst. Sci. 29, 3589–3613.

Jamali, A., Mahdianpari, M., 2022. Swin transformer and deep convolutional neural networks for coastal wetland classification using sentinel-1, sentinel-2, and LiDAR data. Remote. Sens. 14, 359.

Jonkman, S.N., 2005. Global perspectives on loss of human life caused by floods. Nat. Hazards 34, 151–175.

Knoben, W.J.M., Freer, J.E., Woods, R.A., 2019. Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. Hydrol. Earth Syst. Sci. 23, 4323–4331.

Lang, M., Pobanz, K., Renard, B., Renouf, E., Sauquet, E., 2010. Extrapolation of rating curves by hydraulic modelling, with application to flood frequency analysis. Hydrol. Sci. Journal–Journal Des Sci. Hydrol. 55, 883–898.

Li, W., Liu, C., Xu, Y., Niu, C., Li, R., Li, M., Hu, C., Tian, L., 2024. An interpretable hybrid deep learning model for flood forecasting based on transformer and LSTM. J. Hydrol.: Reg. Stud. 54, 101873.

Liao, Y.L., Ho, H.C., 2024. Evaluation of flooding impact with FRI by considering refined model and LID. In: River Flow 2024. CRC Press, pp. 1035–1042.

Lin, Q., Lin, B., Zhang, D., Wu, J., 2022. Web-based prototype system for flood simulation and forecasting based on the HEC-HMS model. Environ. Model. Softw. 158, 105541.

Mahakur, V., Mahakur, V.K., Samantaray, S., Ghose, D.K., 2025. Prediction of runoff at ungauged areas employing interpolation techniques and deep learning algorithm. HydroResearch 8, 265–275.

Bureau of Meteorology, 2024. NSW rainfall and river conditions – flood information for Wagga Wagga Region. https://www.bom.gov.au/nsw/flood/rain_river.shtml. (Accessed 12 June 2025).

Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans. ASABE 50, 885–900.

Moriasi, D.N., Gitau, M.W., Pai, N., Daggupati, P., 2015. Hydrologic and water quality models: Performance measures and evaluation criteria. Trans. ASABE 58, 1763–1785.

Nearing, G., Cohen, D., Dube, V., Gauch, M., Gilon, O., Harrigan, S., Hassidim, A., Klotz, D., Kratzert, F., Metzger, A., et al., 2024. Global prediction of extreme floods in ungauged watersheds. Nature 627, 559–563.

Transport for NSW, 2024. Road segment data (state, regional and local roads) of New South Wales. https://data.nsw.gov.au/data/dataset/86842bb2-65d2-4d4a-a692-d0fd9e35c415. (Accessed 12 June 2025).

NSW Department of Climate Change, Energy, the Environment and Water, 2021. Great soil group (GSG) soil type map of new South Wales (version 4.5). https://data.nsw.gov.au/data/dataset/great-soil-group-gsg-soil-type-map-of-nsw. (Accessed 12 June 2025).

NSW State Emergency Service, 2022. Revised Murrumbidgee River Floodplain Risk Management Study and Plan: Wagga Wagga. Technical Report, NSW State Emergency Service (SES), (Accessed 12 June 2025).

Ocio, D., Le Vine, N., Westerberg, I., Pappenberger, F., Buytaert, W., 2017. The role of rating curve uncertainty in real-time flood forecasting. Water Resour. Res. 53, 4197–4213.

Ouma, Y.O., Omai, L., 2023. Flood susceptibility mapping using image-based 2D-CNN deep learning: Overview and case study application using multiparametric spatial data in data-scarce urban environments. Int. J. Intell. Syst. 2023, 5672401.

Rajib, A., Liu, Z., Merwade, V., Tavakoly, A.A., Follum, M.L., 2020. Towards a large-scale locally relevant flood inundation modeling framework using SWAT and LISFLOOD-FP. J. Hydrol. 581, 124406.

Razali, N., Ismail, S., Mustapha, A., 2020. Machine learning approach for flood risks prediction. IAES Int. J. Artif. Intell. 9, 73.

Rew, R., Davis, G., 1990. NetCDF: An interface for scientific data access. IEEE Comput. Graph. Appl. 10, 76–82.

Samal, P., Patnaik, P., Samantaray, S., Swain, P.C., 2024. Comparative study of reservoir operations using TLBO, PSO and DE optimization techniques: An experiment on the hirakud reservoir, odisha, India. J. Water Clim. Chang. 15, 5309–5331.

Samantaray, S., Sahoo, A., 2024a. Groundwater level prediction using an improved ELM model integrated with hybrid particle swarm optimisation and grey wolf optimisation. Groundw. Sustain. Dev. 26, 101178.

Samantaray, S., Sahoo, A., 2024b. Prediction of flow discharge in Mahanadi River Basin, India, based on novel hybrid SVM approaches. Environ. Dev. Sustain. 26, 18699–18723.

Samantaray, S., Sahoo, A., Agnihotri, A., 2023. Prediction of flood discharge using hybrid PSO-SVM algorithm in barak river basin. MethodsX 10, 102060.

Samantaray, S., Sahoo, A., Yaseen, Z.M., Al-Suwaiyan, M.S., 2025. River discharge prediction based multivariate climatological variables using hybridized long short-term memory with nature inspired algorithm. J. Hydrol. 649, 132453.

Schumann, G., Giustarini, L., Tarpanelli, A., Jarihani, B., Martinis, S., 2023. Flood modeling and prediction using earth observation data. Surv. Geophys. 44, 1553–1578.

Shao, Y., Chen, J., Zhang, T., Yu, T., Chu, S., 2024. Advancing rapid urban flood prediction: A spatiotemporal deep learning approach with uneven rainfall and attention mechanism. J. Hydroinformatics 26, 1409–1424.

Situ, Z., Wang, Q., Teng, S., Feng, W., Chen, G., Zhou, Q., Fu, G., 2024. Improving urban flood prediction using LSTM-DeepLabv3+ and Bayesian optimization with spatiotemporal feature fusion. J. Hydrol. 630, 130743.

Subbarayan, S., Youssef, Y.M., Singh, L., Dąbrowska, D., Alarifi, N., Ramsankaran, R.A.A.J., Visweshwaran, R., Saqr, A.M., 2025. Soil and water assessment tool-based prediction of runoff under scenarios of land use/land cover and climate change across Indian agro-climatic zones: Implications for sustainable development goals. Water 17, 458.

Tavakoli, M., Motlagh, Z.K., Dąbrowska, D., Youssef, Y.M., Đurin, B., Saqr, A.M., 2025. Harnessing AHP and fuzzy scenarios for resilient flood management in arid environments: Challenges and pathways toward sustainability. Water 17, 1276.

Tran, V.N., Kim, T., Xu, D., Tran, H., Le, M.-H., Tran, T.-N.-D., Kim, J., Tran, T.D., Wright, D.B., Restrepo, P., et al., 2025. AI improves the accuracy, reliability, and economic value of continental-scale flood predictions. AGU Adv. 6, e2025AV001678.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. Adv. Neural Inf. Process. Syst. 30.

Wajid, M., Abid, M.K., Raza, A.A., Haroon, M., Mudasar, A.Q., 2024. Flood prediction system using IOT & artificial neural network. VFAST Trans. Softw. Eng. 12, 210–224.

Wang, H., Xu, S., Xu, H., Wu, Z., Wang, T., Ma, C., 2023. Rapid prediction of urban flood based on disaster-breeding environment clustering and Bayesian optimized deep learning model in the coastal city. Sustain. Cities Soc. 99, 104898.

Webb, M., Ltd, A.P., 2008. NSW levee study for emergency management: Detailed urban levee report. https://flooddata.ses.nsw.gov.au/flood-projects/nsw-levee-study-for-emergency-management. (Accessed 12 June 2025).

Willmott, C.J., 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, 79–82.

WMAwater, 2018. Murrumbidgee river at darlington point and environs flood study: Final report. https://www.murrumbidgee.nsw.gov.au/files/assets/public/v/1/publications/town-plans-etc/flood-studies/murrumbidgee-river-flood-study-2018-reduced.pdf. (Accessed 12 June 2025).

Xu, L., Gao, L., 2024. A hybrid surrogate model for real-time coastal urban flood prediction: An application to Macao. J. Hydrol. 642, 131863.

Young, A., Bhattacharya, B., Daniëls, E., Zevenbergen, C., 2025. Integrating WRF forecasts at different scales for pluvial flood forecasting using a rainfall threshold approach and a real-time flood model. J. Hydrol. 656, 132891.

Zeng, Y.-F., Chang, M.-J., Lin, G.-F., 2024. A novel AI-based model for real-time flooding image recognition using super-resolution generative adversarial network. J. Hydrol. 638, 131475.

Zhang, B., Ouyang, C., Cui, P., Xu, Q., Wang, D., Zhang, F., Li, Z., Fan, L., Lovati, M., Liu, Y., et al., 2024. Deep learning for cross-region streamflow and flood forecasting at a global scale. Innov. 5.

Zhang, L., Qin, H., Mao, J., Cao, X., Fu, G., 2023. High temporal resolution urban flood prediction using attention-based LSTM models. J. Hydrol. 620, 129499.

Zhong, M., Xiao, L., Li, X., Mei, Y., Jiang, T., Song, L., Chen, X., 2024. A study on compound flood prediction and inundation simulation under future scenarios in a coastal city. J. Hydrol. 628, 130475.

Zhou, X., Huang, X., Jiang, X., Jiang, J., 2025. Real-time error correction of multiple-hour-ahead flash flood forecasting based on the sliding runoff-rain data and deep learning models. J. Hydrol. 655, 132918.

Zhou, Q., Teng, S., Situ, Z., Liao, X., Feng, J., Chen, G., Zhang, J., Lu, Z., 2023a. A deep-learning-technique-based data-driven model for accurate and rapid flood predictions in temporal and spatial dimensions. Hydrol. Earth Syst. Sci. 27, 1791–1808.

Zhou, Y., Wu, Z., Xu, H., Yan, D., Jiang, M., Zhang, X., Wang, H., 2023b. Adaptive selection and optimal combination scheme of candidate models for real-time integrated prediction of urban flood. J. Hydrol. 626, 130152.