Urban Flood-Depth Forecasting with Autoregressive Surrogates: A Case Study of Indian Cities

Pallavi Tyagi*[1]
tyagipallavi093@gmail.com
Vishal Dubey*[2]
vishaldubey0026@gmail.com

[1] Bengaluru, India
[2] Hyderabad, India

Abstract

Accurate, fast flood-depth forecasts are essential for urban risk management, yet physics-based 2D hydraulics are often too slow for rapid, large-area use. We develop data-driven neural surrogates that learn spatio-temporal flood depth from multi-region events: an autoregressive feed-forward network (FNN2D-AR) that rolls short-horizon predictions forward, and a space-time convolutional baseline (FNN3D) producing fixed multi-step outputs. Using FloodCastBench sequences, we test cross-country generalization and perform external validations for the 2005 Mumbai flood and the 2015 Chennai flood. Overall, the ML model produces flood-depth forecasts with promising cross-region transfer, supporting rapid mapping and planning. Code link is also shared: Code

1 Introduction and Related Work

Urban flooding increasingly threatens lives, infrastructure, and equity in growing cities, making accurate, fast flood-depth maps essential for risk management and planning. Physics-based hydraulics for flood depth prediction are accurate but slow and Machine Learning (ML) surrogates are fast yet often struggle to generalize. However, a growing line of work in the broader urban flooding domain investigating cross-region transfer reports that pretraining on large, diverse source regions followed by fine-tuning on a target region yields consistent gains on the target domain [39, 43, 47]. This pattern aligns with broader evidence on model generalization under covariate shift, where heterogeneity in hydrology, built form, and rainfall regimes can degrade out-of-domain performance unless the model is adapted [40]. Together, these results motivate multi-region training and lightweight adaptation protocols for operational deployment in new cities. Building on these evidence from scalable ML forecasting and adaptation [6, 24, 26, 30, 34, 36, 37], we learn spatio-temporal depth from multi-region events and test cross-country transfer with light fine-tuning and external validation.

Many recent studies train machine-learning surrogates directly on outputs from established hydrodynamic simulators such as Delft3D-FM, LISFLOOD-FP, MIKE FLOOD, and HEC-RAS, using solver-generated depth fields as labels for deep or tree-based models [2, 4, 11,

*Equal contribution. This work is independent research.

[14, 21, 33, 49]. Experimental designs are typically tied to specific topographies and forcings, including synthetic or random terrains for breach scenarios, fixed-bathymetry river reaches, and dense urban cores driven by designed or synthetic storms, and these surrogates generally achieve high fidelity with large computational savings; some studies focus on maximum-depth mapping from DEM and land-use features [2, 4, 11, 14, 16, 21, 33, 49]. However, cross-region transferability is limited or insufficiently assessed, performance can deteriorate under feature shift, and many evaluations remain city or catchment specific [22]. A smaller line of work targets point-scale prediction at gauges rather than maps: LSTM models that fuse recent depth with short-lead rainfall forecasts retain skill over multi-step rollouts while tree-based models with SHAP provide site-level factor rankings and strong point-fit [15, 23]. In parallel, purely data-driven nowcasting predicts short-lead flood depth or extent maps using only recent observations or proxies, showing utility without external drivers [25, 46]; coastal datasets have been assembled to expose models to surge- and tide-affected patterns during training [46]. These approaches are physics-agnostic and, at gauges, do not yield citywide, spatially consistent depth fields; while observation-only nowcasting reduces input burden, its skill can be sensitive to regime changes not represented in the history, and transferability across storm regimes and cities remains incompletely tested [22].

Apart from data-driven surrogates, several studies embed physics to improve credibility and transfer. They inject physics via simulator-trained surrogates, multi-fidelity refinements of coarse hydraulics, drainage-supervised overflow routing to 2-D inundation, SAR plus DEM water-surface reconstruction, and physics-informed solvers embedded in forecasting systems, including GeoPINS within FloodCast [5, 7, 13, 20, 29, 38]. Across national and urban contexts these methods report high fidelity and operational speed, often improving spatial coherence and interpretability. Generalizability remains mixed; many approaches depend on calibrated local simulators, detailed drainage inventories, or coarse grids, and transfer to new cities often requires retraining [5, 7, 13, 20, 29, 38].

Generalizability and transferability remain central concerns in flood-depth prediction, although a growing set of studies seeks to address them. A volume-conserving cGAN surrogate trained on 10 catchments transfers to 5 unseen with large speed-ups but underestimates channel depths in small-peak events [8]; in Berlin, Random Forest is strong in-domain yet overfits, whereas U-Net generalizes better and benefits from $\sim 10\%$ target fine-tuning [41]; for the Fitzroy River, a single MIKE21-trained surrogate over 224 tiles yields $57\times$ speed-ups but omits local rainfall and is shown for one basin [45]. Overall, physics guidance or transfer learning can enable cross-area transfer, but reliance on simulator labels, local calibration, and retraining under feature shift persists.

In the related domain of numerical weather prediction, recent work shows that pretraining on diverse climates and making uncertainty explicit can improve model generalization. Deterministic and probabilistic ML forecasters have been developed for extreme precipitation and related fields at global to regional scales[6, 24, 26, 30, 34]. These studies document the benefits of scaling data and model capacity, provide uncertainty estimates, and highlight the value of standardised evaluation. Benchmark suites such as WeatherBench/WeatherBench 2 have catalysed progress by fixing data, splits, and metrics [36, 37].

Beyond methods, progress on cross-region transfer depends on heterogeneous datasets that capture varied hydrology, built form, measurement conditions, and labeling sources [12, 19, 27, 28, 48]. Multiple challenges motivate this need, including detection ambiguities in urban remote sensing and ancillary data and limited ground truth in dense cities, which complicate training and evaluation [27]. In response, several initiatives curate flood datasets spanning methods and modalities, including repositories derived from hydrodynamic simulations, SAR-

validated inundation products, and global floodplain layers [12, 19, 28, 48]. These collections differ in spatial and temporal resolution, label provenance (physics vs. observation), and event typology (pluvial, fluvial, coastal), and this diversity is valuable for studying generalization and for stress-testing models across hydrologic regimes.

This study addresses the generalization gap in flood depth prediction and proposes ML method which learns spatio-temporal flood depth from multi-region events. We employ a Fourier Neural Operator Network (FNN), an operator-learning approach for evolving fields, trained on FloodCastBench [48]. This dataset covers regions like Australia, the United Kingdom, Mozambique, and Pakistan. We studied potential of data driven ML model trained on multiple countries to generalize to an unseen country and to new Indian cities. We train on events from Australia, the United Kingdom, and Mozambique; evaluate on a held-out country (Pakistan); and conduct an external validation in Mumbai for the 2005 flood and Chennai for the 2015 flood. Fast, transferable citywide depth forecasts with light adaptation will enable earlier, targeted warnings, and more equitable allocation of adaptation resources in data-scarce cities.

2 Methodology

Problem formulation. Let $D_t \in \mathbb{R}^{H\times W}$ denote the flood-depth field at time t on a fixed grid (5-minute sampling). Let $S = \{S^{\mathrm{DEM}}, S^{\mathrm{LULC}}\}$ be static rasters (digital elevation and land-use/roughness proxies), and let R_t be the rainfall intensity field. Given an initial condition D_t, static covariates S, spatial/temporal coordinates, and a sequence of future rainfall fields $R_{t+1:t+T}$, the task is to forecast future depth maps $D_{t+1:t+T}$.

FNN2D-AR (autoregressive) model. The FNN2D-AR consumes a stack comprising the initial depth D_t, static rasters S, coordinates, the lead-time channel ℓ, and rainfall covariates. The network outputs either a one-step prediction $\hat{D}_{t+1}$ or a short block $\hat{D}_{t+1:t+T}$ autoregressively [30, 32, 35] as shown in Figure 1. At inference, predictions are rolled out autoregressively:

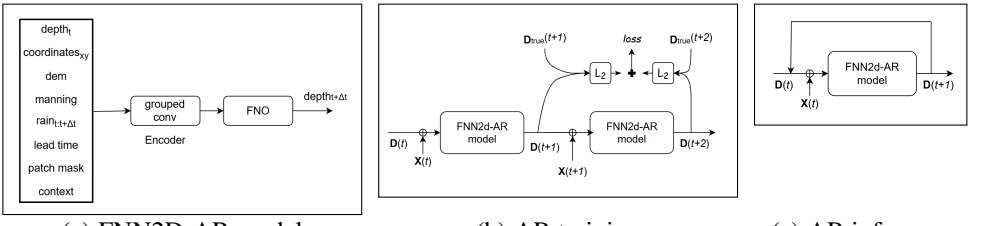
$$\tilde{D}_{t+\tau} = \text{FNN2D-AR}\big(\tilde{D}_{t+\tau-1}, S, R_{t+\tau}, \ell\big), \qquad \tau = 1,2,\dots \tag{1}$$

FNN3D (space–time baseline) model. The FNN3D applies 3D convolutions over a fixed temporal window. Following the required setting, it takes the *first* time-step depth map D_t and *replicates* it along the time axis to form a length-T depth tensor $\underbrace{[D_t, \dots, D_t]}_{T \text{ copies}}$. This replicated sequence, concatenated with static rasters S, the lead-time channel ℓ and the future rainfall sequence $R_{t+1:t+T}$, is mapped directly to the multi-step output $\hat{D}_{t+1:t+T}$ in a single forward pass. The FNN3D does not consume context channels (e.g., masks or coarse summaries).

2.1 Model architecture

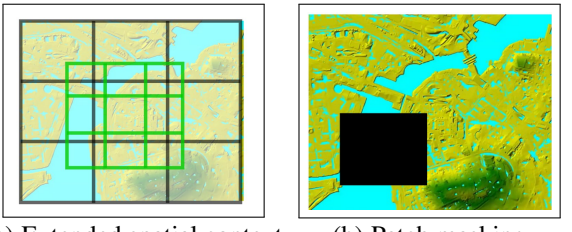
Encoder block. We adopt a lightweight encoder that expands per-pixel feature dimensionality using grouped 1×1 convolutions [3, 50]. Concretely, let $X \in \mathbb{R}^{H\times W\times C_{\text{in}}}$ denote the input stack (depth, DEM, LULC, rainfall, coordinates, etc.). The encoder applies a 1×1 convolution with groups equal to C_{in}, producing $Z \in \mathbb{R}^{H\times W\times C_{\text{latent}}}$ as shown in Figure 1. This operation increases channel capacity. Only used for FNN2D-AR.

Square patch masking. To improve robustness to missing or corrupted inputs (e.g., sensor gaps, occlusions), we apply square patch masking to the input stack during training as shown in Figure 2, following recent work in geoscience imputation and nowcasting [1, 16, 46]. We replace masked regions by a learnable token and token is optimised jointly with model parameters. Only used for FNN2D-AR.



(a) FNN2D-AR model (b) AR training (c) AR inference

Figure 1: (a) shows FNN2D-AR which takes depth, DEM, manning raster etc. as input channels followed by Encoder followed by Fourier Neural Operator (FNO) block. (b) and (c) shows autoregressive training and inference. D(t) is model predicted flood depth and X(t) is input channels like DEM, manning, rainfall etc.



(a) Extended spatial context (b) Patch masking

Figure 2: FNN2D-AR model input components.

Extended spatial context. Following large-context conditioning in weather and flood forecasting [9], each target patch is augmented with a larger surrounding window (context) from the same time step as shown in Figure 2. The context window is downsampled to the patch resolution and concatenated channel-wise. Only used for FNN2D-AR.

Multi-resolution temporal sampling. We train at multiple temporal resolutions $\Delta t \in \{15, 30, 60\}$ minutes, randomly sampling one Δt per sequence during training [30]. We encode the selected temporal resolution via a scalar *lead-time channel* appended to the inputs, analogous to conditioning used in scalable nowcasting [9].

2.2 Training procedure and Dataset

Dataset and splits We use *FloodCastBench* [48]. Within each training country we adopt a chronological split: the first 70% of time steps for training and the last 30% for validation (to avoid leakage across storm phases). Sequences are tiled into fixed-size patches with random

spatial crops and temporal windows, and sampling is balanced across regions. All inputs are reprojected to a common grid.

Curriculum over horizon. FNN2D-AR is trained progressively: first to predict a single step (T=1), then fine-tuned for T=8, and finally for T=20 as shown in Figure 1. The FNN3D baseline is trained directly for a T=20-step output window.

Inference rollout. Inference is fully autoregressive for FNN2D-AR as shown in Figure 1 and single-pass for FNN3D.

Loss and optimization. The training objective is a *batch-wise relative p-norm loss* computed over all elements of each example (all pixels and, when applicable, all lead times). Let $x, y \in \mathbb{R}^{B\times\cdots}$ denote predicted and target tensors in a batch of size B, and let $\mathrm{vec}(\cdot)$ flatten the non-batch dimensions. The loss is

$$\mathcal{L}_{\text{rel-}p}(x,y) \;=\; \frac{1}{B}\sum_{b=1}^{B} \frac{\left\|\mathrm{vec}(x_b)-\mathrm{vec}(y_b)\right\|_p}{\left\|\mathrm{vec}(y_b)\right\|_p+\varepsilon}, \tag{2}$$

where p is the norm order (we use p=2), and ε is a small constant for numerical stability. Early stopping on validation loss is used for model selection. All training experiments are run on a single NVIDIA A100 (40 GB) GPU.

3 Results

Runtime and memory efficiency. We benchmark inference on an NVIDIA T4 (15 GB VRAM) across horizons $TS \in \{20, 40, 80\}$, as shown in Table 1. These measurements support the choice of the compact autoregressive model for long-horizon deployment under tight GPU memory budgets.

Table 1: Inference latency and memory footprint on NVIDIA T4 GPU (15 GB VRAM). Latency in milliseconds, memory usage in MB, for each TS time-steps

Model	#Params	TS=20		TS=40		TS=80	
		Latency (ms)	Mem (MB)	Latency (ms)	Mem (MB)	Latency (ms)	Mem (MB)
FNN3D	10,427,717	91.36	1454.19	172.87	2759.26	336.97	5269.06
FNN2D-AR	7,859,762	269.81	389.10	533.74	389.20	1048.11	389.20
FNN2D-AR (small)	**576,178**	**83.01**	**209.38**	**165.97**	**209.38**	**331.59**	**209.38**

Cross-country generalization on Pakistan. We first evaluate cross-country transfer by training on UK/Australia/Mozambique and holding out Pakistan for testing. Table 2 in the paper reports scores at a forecast horizon of TS=20. These results suggest that compact autoregressive models can balance error accumulation and generalization more effectively at this horizon. Figure 3 contrasts depth predictions and error maps for all three models on representative Pakistan scenes. Visual inspection indicates that the compact autoregressive model reduces spatial residuals relative to the other variants. While all models capture the gross inundation footprint, errors concentrate near wet/dry fronts, which is consistent with the higher sensitivity of those areas to temporal integration and rainfall timing.

Table 2: Evaluation on Pakistan, inference resolution@224×224, TS=20 (each 60 minute gap). Evaluation metrics include root mean square error (RMSE, m), Nash-Sutcliffe efficiency (NSE), Pearson correlation, and Critical Success Index (CSI) at (0.01 m, 0.001 m).

Model	RMSE ↓	NSE ↑	Pearson r ↑	CSI@0.001 m ↑	CSI@0.01 m ↑
FNN3D	0.373	0.569	**0.895**	0.963	**0.901**
FNN2D-AR	0.375	0.599	0.860	0.958	0.772
FNN2D-AR (small)	**0.333**	**0.711**	0.863	**0.975**	0.873

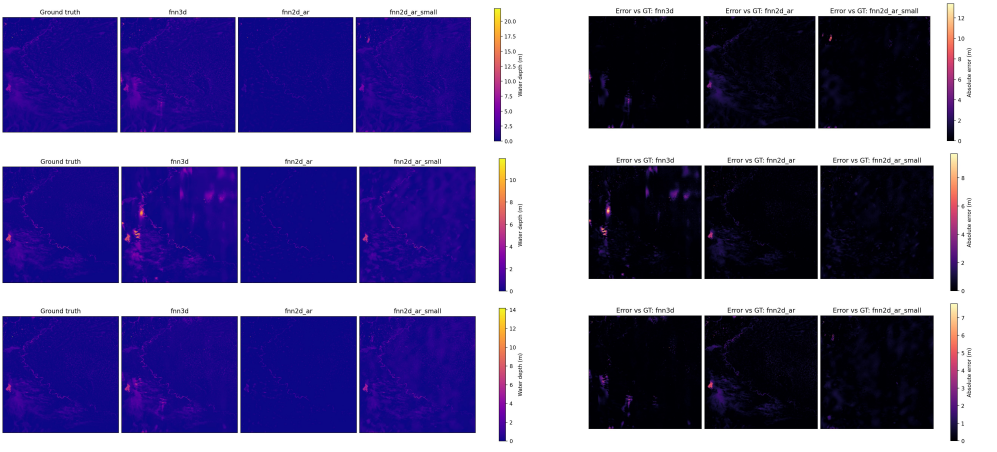
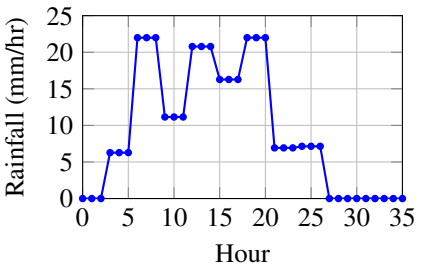


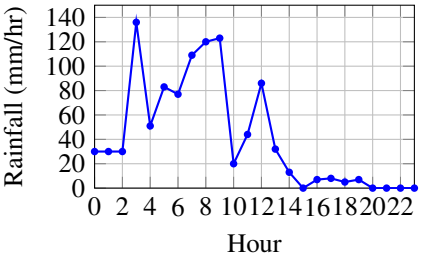
Figure 3: Comparison of paired results for three models on Pakistan flood. Each row shows two corresponding images for the same case: first column showing Ground Truth (GT) and model predicted depth maps and second column showing error map of three models w.r.t. GT depth map. Here, TS=20 (each 60 minute gap).

External city-level validation: Mumbai 2005 flood. Mumbai, a coastal city, experienced catastrophic flooding on 26 July 2005 with severe loss of life and disruption [31]. The Mithi River routes overflows from the Vihar and Powai lakes to the Arabian Sea at Mahim Creek; tidal backwater, environmental degradation, channel encroachment, and a flattened lower gradient constrained outflow and amplified overbank inundation [10, 50]. For this study, we evaluate transfer on the 26–27 July 2005 Mithi flood using SRTMGL1 v003 (30 m) terrain, the Vihar hourly hyetograph (IST; 26 Jul 12:00 to 27 Jul) from the official report [10], and a 30 m GLC-FCS30D collection in Earth Engine land-cover map for HEC-HMS losses. Inflows are generated with HEC-HMS (SCS-CN, SCS-UH, kinematic wave) following [18, 44, 50, 51] and routed in a 2D unsteady HEC-RAS domain to produce reference depth fields [42]. We have run a pre-run in HEC-RAS with 15% of peak hydrograph discharge for 24 hours to generate initial conditions. HEC-RAS is simulated for 24 hour main event rainfall as in Figure 4 and this generated depth maps at 15 minutes interval, creating the dataset. Our ML Model is Finetuned (FT) with first 35% of dataset and evaluated on remaining 65% dataset. Model is FT with very low learning rate. Table 3 shows performance comparison of model with and without FT. Figure 5 shows model results with and without FT.

External city-level validation: Chennai 2015 flood. To examine transfer beyond the dataset's countries, we conduct a qualitative validation on the 1-2 December 2015 Chennai

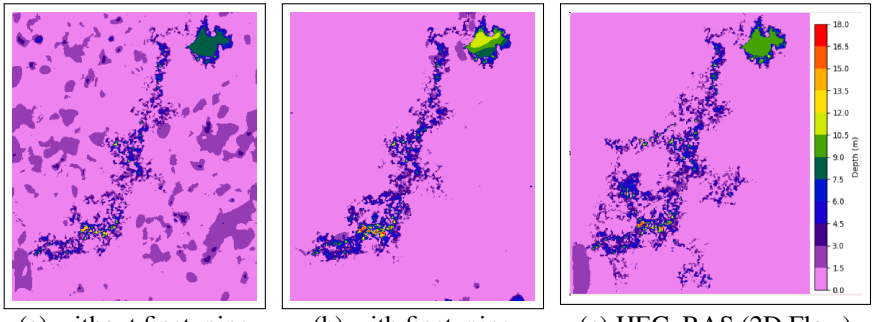


(a) Chennai December 2015 rainfall



(b) Mumbai July 2005 rainfall

Figure 4: Flood regions rainfalls



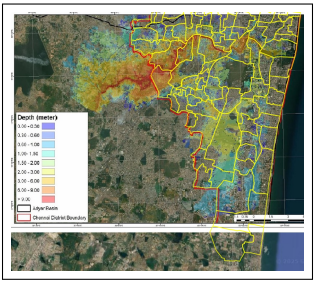
(a) without finetuning (b) with finetuning (c) HEC–RAS (2D Flow)

Figure 5: Comparison of FNN2D-AR (small) model predictions of flood depth ((a) without finetuning and (b) with finetuning) with HEC-RAS (considered Ground Truth here) generated flood depth, for Mumbai 2005 flood. Initial flood depth is same for all and all results are maximum flood depth.

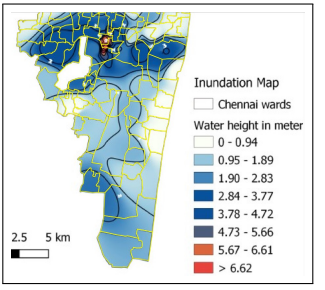
Table 3: Evaluation on Mumbai 2005 floods for FNN2D-AR(small). M1 = model without finetuning (FT); M2 = model with FT, for increasing TS time-steps (each 30 minute gap).

Eval. metric	TS=8		TS=16		TS=29	
	M1	M2	M1	M2	M1	M2
RMSE ↓	0.70	**0.37**	0.89	**0.53**	1.07	**0.70**
NSE ↑	0.81	**0.95**	0.70	**0.89**	0.56	**0.81**
Pearson r ↑	0.92	**0.97**	0.87	**0.95**	0.78	**0.90**
CSI@0.01 m ↑	0.18	**0.25**	0.19	**0.26**	0.18	**0.27**

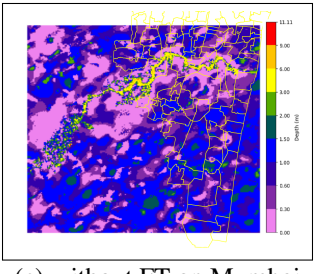
flood in Adyar river basin. In this study, Elevation comes from the SRTMGL1 v003 DEM (30 m; Google Earth Engine asset USGS/SRTMGL1_003). Land use/land cover is from the 30 m GLC-FCS30D collection in Earth Engine; we use the 2015 layer mosaicked over the Adyar basin. Event precipitation is TRMM 3B42 (TMPA, V7; 3 -hourly, 0.25°, mm h^{-1}; Earth Engine asset TRMM/3B42). Subbasin, Junctions, boundary condition lines etc., are created as per [42]. We have run a pre-run in HEC-RAS with 10% of peak hydrograph discharge for 72 hours to generate initial flood depth conditions following [42]. FNN2D-AR (small) which is FT on Mumbai is evaluated for cross-city predictions for rainfall event as in Figure 4. Figure 6 compares the FNN2D-AR (small) depth map against two independent references



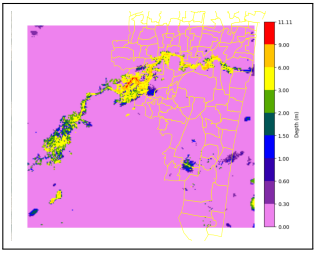
(a) Flood depth from [42]



(b) Flood depth from [17]



(c) without FT on Mumbai



(d) with FT on Mumbai

Figure 6: Comparison of FNN2D-AR (small) model predictions flood depth with [17, 42], for Chennai 2015 flood. Yellow polygons are drawn for easy visual comparison.

[17, 42] for the same event. As shown in Figure 6, the qualitative alignment of model without FT indicates that the learned surrogate can transfer to a new Indian coastal city without local retraining. But, with local retraining of model, precision and accuracy of model flood depth prediction improves.

4 Discussion

Our results show the effectiveness of proposed methodology of AR training, Encoder block, extended spatial context, patch masking and multi-resolution temporal sampling as shown in Table 2.

We find that compact autoregressive models generalize most reliably: the small FNN2D-AR attained the best aggregate scores on Pakistan while maintaining a favourable latency-memory profile, supporting its suitability for long horizons under constrained GPUs. Larger AR variant and FNN3D seems to overfit to training regions and predicting more aggresive and sharp depth changes while FNN2D-AR (small) is conservative in flood depth prediction, leading to better generalization as shown in Figure 3. Additionally, runtime and memory scale differently across designs with iterative rollout keeping memory nearly constant with horizon (beneficial for deployment), whereas single-pass multi-step decoding increases memory substantially.

We performed Hydrodynamic flood simulation of Mumbai 2005 floods for Mithi river basin using HEC-HMS and HEC-RAS. We evaluated model generalization on these simulation depth maps and observed that model accuracy is decent even without India specific finetuning. However, this model predicts false flooding regions. To evaluate efficacy of FT on Mumbai

region with small dataset, we finetuned AR model on flood simulation maps and observed that model accuracy improves further as can be seen in Table 3. Generated depth maps of FT model aligns well with simulation depth maps as can be seen in Figure 5. As with autoregressive models, performance drops with more inference timesteps.

We evaluated efficacy of model finetuned on Mumbai to another Indian city. We simulated Chennai 2015 flood event and used this study [42] as reference to generate initial flood depth. As can be observed in Figure 6 (c), model which is not FT on Mumbai gives similar flood depth to Figure 6 (a) in Adyar river channel and other regions (inside yellow polygons) but also gives false flood depth regions (outside yellow polygons). As shown in Figure 6 (d), Model finetuned on Mumbai gives even more accurate flood depth in Adyar river channel compared to Figure 6 (a). Also, this finetuned model predictions aligns with Figure 6 (b) (inside yellow polygons) and with Figure 6 (a) (outside yellow polygons).

In this study we use DEM, rainfall sequences, and a surface-roughness layer. However, the current models do not explicitly couple drainage or tide and rely on physics-generated training labels; consequently, performance is sensitive to the quality of rainfall forcing and the representativeness of the training hydrologic regimes. Moreover, our evaluation focuses on one held-out country and two Indian city events; broader assessments across different climate zones and urban infrastructure (drainage networks, land use) would further strengthen generalization claims. Finally, we do not yet condition on catchment descriptors (e.g., slope, soils, land cover, drainage density) or on climatology; however, incorporating these may improve robustness under covariate shift.

5 Conclusion

We develop fast, data-driven surrogates for city-scale flood-depth forecasting from multi-region events: FNN2D-AR and FNN3D. We test cross-country generalization on Pakistan and validate externally on Chennai 2015 (qualitative) and Mumbai 2005 (quantitative vs HEC-RAS). Compact autoregressive models generalize best: the small FNN2D-AR achieves top aggregate scores on Pakistan with favorable latency and memory. Mumbai results confirm transfer and benefits from small, region-specific fine-tuning, and Chennai shows reproduction of large-scale inundation, indicating transfer beyond training countries. By reducing dependence on local simulators and extensive retraining, the system expands deployability in data-scarce cities and accelerates preparedness. We aim to extend this study along several directions: (i) conditioning the surrogate on documented catchment descriptors and climatology by adding them as explicit inputs and assessing their effect on cross-city transfer; (ii) exploring coupling of drainage and tidal boundary information by incorporating drainage exchange and downstream stage time series into the forecasting pipeline; (iii) broadening held-out evaluations across different climate zones and urban infrastructure by adding additional countries and cities, reporting both zero-shot and light fine-tuning results under the same accuracy and efficiency metrics; and (iv) investigating light physics cues (mass and flux consistency terms and drainage priors) to stabilise out-of-distribution behaviour, applying few-shot domain adaptation for rapid local tuning, and exploring assimilation of event imagery such as SAR for real-time correction.

References

[1] Marcin Andrychowicz, Lasse Espeholt, Di Li, Samier Merchant, Alexander Merose, Fred Zyda, Shreya Agrawal, and Nal Kalchbrenner. Deep learning for day forecasts from sparse observations. *arXiv preprint arXiv:2306.06079*, 2023.

[2] Roberto Bentivoglio, Elvin Isufi, Sebastiaan Nicolas Jonkman, and Riccardo Taormina. Rapid spatio-temporal flood modelling via hydraulics-based graph neural networks. *Hydrology and Earth System Sciences*, 27(23):4227–4246, 2023.

[3] Boris Bonev, Thorsten Kurth, Ankur Mahesh, Mauro Bisson, Jean Kossaifi, Karthik Kashinath, Anima Anandkumar, William D Collins, Michael S Pritchard, and Alexander Keller. Fourcastnet 3: A geometric approach to probabilistic machine-learning weather forecasting at scale. *arXiv preprint arXiv:2507.12144*, 2025.

[4] Priyanka Chaudhary, João P Leitão, Konrad Schindler, and Jan Dirk Wegner. Flood water depth prediction with convolutional temporal attention networks. *Water*, 16(9): 1286, 2024.

[5] Lakshmi Amani Chimata, Suresh Babu Anuvala Setty Venkata, Shashi Vardhan Reddy Patlolla, Durga Rao Korada Hari Venkata, Sreenivas Kandrika, and Prakash Chauhan. Automated rapid estimation of flood depth using a digital elevation model and earth observation satellite (eos-04)-derived flood inundation. *Natural Hazards and Earth System Sciences*, 25(7):2455–2472, 2025.

[6] Guillaume Couairon, Christian Lessig, Anastase Charantonis, and Claire Monteleoni. Archesweather: An efficient ai weather forecasting model at 1.5° resolution. *arXiv e-prints*, pages arXiv–2405, 2024.

[7] Peng Cui, Nazir Ahmed Bazai, Zou Qiang, Wang Jiao, Wang Yan, Qingsong Xu, Lei Yu, and Zhang Bo. Flood risk assessment with machine learning: insights from the 2022 pakistan mega-flood and climate adaptation strategies. *npj Natural Hazards*, 2(1):42, 2025.

[8] Cesar AF Do Lago, Marcio H Giacomoni, Roberto Bentivoglio, Riccardo Taormina, Marcus N Gomes Junior, and Eduardo M Mendiondo. Generalizing rapid flood predictions to unseen urban catchments with conditional generative adversarial networks. *Journal of Hydrology*, 618:129276, 2023.

[9] Lasse Espeholt, Shreya Agrawal, Casper Sønderby, Manoj Kumar, Jonathan Heek, Carla Bromberg, Cenk Gazen, Rob Carver, Marcin Andrychowicz, Jason Hickey, et al. Deep learning for twelve hour precipitation forecasts. *Nature communications*, 13(1):5145, 2022.

[10] Fact Finding Committee on Mumbai Floods. Fact finding committee on mumbai floods: Final report, volume i, March 2006. URL https://www.scribd.com/document/293669664/Fact-Finding-Committee-on-Mumbai-Floods-Vol1. Chair: Madhavrao Chitale.

[11] Mohamed M Fathi, Zihan Liu, Anjali M Fernandes, Michael T Hren, Dennis O Terry, C Nataraj, and Virginia Smith. Spatiotemporal flood depth and velocity dynamics using a convolutional neural network within a sequential deep-learning framework. *Environmental Modelling & Software*, 185:106307, 2025.

[12] Mulham Fawakherji, Jeffrey Blay, Matilda Anokye, Leila Hashemi-Beni, and Jennifer Dorton. Deepflood for inundated vegetation high-resolution dataset for accurate flood mapping and segmentation. *Scientific Data*, 12(1):271, 2025.

[13] Niels Fraehr, Quan J Wang, Wenyan Wu, and Rory Nathan. Assessment of surrogate models for flood inundation: The physics-guided lsg model vs. state-of-the-art machine learning models. *Water Research*, 252:121202, 2024.

[14] Francisco Haces-Garcia, Natalya Ross, Craig L Glennie, Hanadi S Rifai, Vedhus Hoskere, and Nima Ekhtari. Rapid 2d hydrodynamic flood modeling using deep learning surrogates. *Journal of Hydrology*, 651:132561, 2025.

[15] Jing Huang, Yonghang Hong, and Dianchen Sun. Urban flood depth prediction using an improved lstm model incorporating precipitation forecasting. *Natural Hazards*, 121(7): 8305–8326, 2025.

[16] Areg Karapetyan, Aaron Chung Hin Chow, and Samer Madanat. Deep vision-based framework for coastal flood prediction under climate change impacts and shoreline adaptations. *arXiv preprint arXiv:2406.15451*, 2024.

[17] Dhivya Karmegam, Sivakumar Ramamoorthy, and Bagavandas Mappillairaju. Near real time flood inundation mapping using social media data as an information source: a case study of 2015 chennai flood. *Geoenvironmental Disasters*, 8(1):25, 2021.

[18] Mujiburrehman Khan. Significance of mangroves in flood protection of coastal area: A case study of mithi river, mumbai, india. *World Journal of Environmental Biosciences*, 3(2):97–108, 2014.

[19] M. Le Gal, T. Fernández-Montblanc, E. Duo, J. Montes Perez, P. Cabrita, P. Souto Ceccon, V. Gastal, P. Ciavola, and C. Armaroli. A new european coastal flood database for low–medium intensity events. *Natural Hazards and Earth System Sciences*, 23 (11):3585–3602, 2023. doi: 10.5194/nhess-23-3585-2023. URL https://nhess.copernicus.org/articles/23/3585/2023/.

[20] Cheng-Chun Lee, Lipai Huang, Federico Antolini, Matthew Garcia, Andrew Juan, Samuel D Brody, and Ali Mostafavi. Predicting peak inundation depths with a physics informed machine learning model. *Scientific Reports*, 14(1):14826, 2024.

[21] Jingru Li, Guiying Pan, Yangyu Chen, Xiaoling Wang, Peizhi Huang, Li Zhang, and Haijun Zhou. Rapid-mapping maximum water depth map of urban flood using a highly adaptable machine learning based model. *Journal of Flood Risk Management*, 18(3): e70095, 2025.

[22] Bo Liu, Yingbing Li, Minyuan Ma, and Bojun Mao. A comprehensive review of machine learning approaches for flood depth estimation: Liu et al. machine learning approaches for flood depth estimation. *International Journal of Disaster Risk Science*, pages 1–13, 2025.

[23] Yuan Liu, Hongfa Wang, Xinjian Guan, Yu Meng, and Hongshi Xu. Urban flood depth prediction and visualization based on the xgboost-shap model. *Water Resources Management*, 39(3):1353–1375, 2025.

[24] Ignacio Lopez-Gomez, Amy McGovern, Shreya Agrawal, and Jason Hickey. Global extreme heat forecasting using neural weather models. *Artificial Intelligence for the Earth Systems*, 2(1):e220035, 2023.

[25] Samuel Chege Maina and Eric Wanjau. Towards flood extent forecasting: Evaluating a weather foundation model and u-net for flood forecasting. In *Climate Change AI*, April 2025. URL https://www.microsoft.com/en-us/research/publication/towards-flood-extent-forecasting-evaluating-a-weather-foundat

[26] Cristian Meo, Ankush Roy, Mircea Lică, Junzhe Yin, Zeineb Bou Che, Yanbo Wang, Ruben Imhoff, Remko Uijlenhoet, and Justin Dauwels. Extreme precipitation nowcasting using transformer-based generative models. *arXiv preprint arXiv:2403.03929*, 2024.

[27] Amit Misra, Kevin White, Simone Fobi Nsutezo, William Straka III, and Juan Lavista. Mapping global floods with 10 years of satellite radar data. *Nature Communications*, 16 (1):5762, 2025.

[28] Fernando Nardi, A Annis, Giuliano Di Baldassarre, ER Vivoni, and Salvatore Grimaldi. Gfplain250m, a global high-resolution dataset of earth's floodplains. *Scientific data*, 6 (1):1–6, 2019.

[29] Huu Duy Nguyen, Dinh Kha Dang, Nhu Y Nguyen, Chien Pham Van, Thi Thao Van Nguyen, Quoc-Huy Nguyen, Xuan Linh Nguyen, Le Tuan Pham, Viet Thanh Pham, and Quang-Thanh Bui. Integration of machine learning and hydrodynamic modeling to solve the extrapolation problem in flood depth estimation. *Journal of Water and Climate Change*, 15(1):284–304, 2024.

[30] Tung Nguyen, Rohan Shah, Hritik Bansal, Troy Arcomano, Sandeep Madireddy, Romit Maulik, Veerabhadra Kotamarthi, Ian Foster, and Aditya Grover. Scaling transformers for skillful and reliable medium-range weather forecasting. In *ICLR 2024 Workshop on AI4DifferentialEquations In Science*, 2024.

[31] Archana Mahesh Patankar. The exposure, vulnerability, and ability to respond of poor households to recurrent floods in mumbai. *World Bank Policy Research Working Paper*, (7481), 2015.

[32] Jaideep Pathak, Shashank Subramanian, Peter Harrington, Sanjeev Raja, Ashesh Chattopadhyay, Morteza Mardani, Thorsten Kurth, David Hall, Zongyi Li, Kamyar Azizzadenesheli, et al. Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators. *arXiv preprint arXiv:2202.11214*, 2022.

[33] Matteo Pianforini, Susanna Dazzi, Andrea Pilzer, Renato Vacondio, et al. Real-time flood maps forecasting for dam-break scenarios with a transformer-based deep learning model. *Journal of Hydrology*, 635:131169, 2024.

[34] Ilan Price, Alvaro Sanchez-Gonzalez, Ferran Alet, Tom R Andersson, Andrew El-Kadi, Dominic Masters, Timo Ewalds, Jacklynn Stott, Shakir Mohamed, Peter Battaglia, et al. Probabilistic weather forecasting with machine learning. *Nature*, 637(8044):84–90, 2025.

[35] Vivek Ramavajjala. Heal-vit: Vision transformers on a spherical mesh for medium-range weather forecasting. *arXiv preprint arXiv:2403.17016*, 2024.

[36] Stephan Rasp, Peter D Dueben, Sebastian Scher, Jonathan A Weyn, Soukayna Mouatadid, and Nils Thuerey. Weatherbench: a benchmark data set for data-driven weather forecasting. *Journal of Advances in Modeling Earth Systems*, 12(11):e2020MS002203, 2020.

[37] Stephan Rasp, Stephan Hoyer, Alexander Merose, Ian Langmore, Peter Battaglia, Tyler Russell, Alvaro Sanchez-Gonzalez, Vivian Yang, Rob Carver, Shreya Agrawal, et al. Weatherbench 2: A benchmark for the next generation of data-driven global weather models. *Journal of Advances in Modeling Earth Systems*, 16(6):e2023MS004019, 2024.

[38] Hancheng Ren, Bo Pang, Gang Zhao, Haijun Yu, Peinan Tian, and Chenran Xie. Incorporating dynamic drainage supervision into deep learning for accurate real-time flood simulation in urban areas. *Water Research*, 270:122816, 2025.

[39] Emil Ryd and Grey Nearing. Fine flood forecasts: Incorporating local data into global models through fine-tuning. *arXiv preprint arXiv:2504.12559*, 2025.

[40] Bilal Sardar and Lakshmi Babu Saheer. Flood detection modeling: Leveraging the sen1flood11 dataset for the rio colima river.

[41] Omar Seleem, Georgy Ayzel, Axel Bronstert, and Maik Heistermann. Transferability of data-driven models to predict urban pluvial flood water depth in berlin, germany. *Natural Hazards and Earth System Sciences Discussions*, 2022:1–23, 2022.

[42] Sunil Shivakumar, Monisha Shanmuganathan, Ashantha Goonetilleke, Darryl Day, Archana Sarkar, Dharmappa Hagare, Basant Maheshwari, Ricky Spencer, Matadadoddi Nanjundegowda Thimmegowda, Damodhara Rao Mailapalli, et al. Managing floods in chennai city as part of situation understanding and improvement project. *World Water Policy*, 9(3):349–370, 2023.

[43] Helen Tamura-Wicks, Geoffrey Dawson, Andrew Taylor, Chris Dearden, Anne Jones, and Paolo Fraccaro. Earth observation foundation models for region-specific flood segmentation. In *International Conference on Learning Representations*, 2025.

[44] Pallavi Tyagi. Flood risk, coastal megacities, and urban poor: assessing the future urban flood risk in the h/e ward of mumbai. *Journal of Urban and Environmental Engineering*, 14(2):191–203, 2020.

[45] Lachlan Tychsen-Smith, Mohammad Ali Armin, and Fazlul Karim. Towards non-region specific large-scale inundation modelling with machine learning methods. *Water*, 16 (16):2263, 2024.

[46] Brandon Victor, Mathilde Letard, Peter Naylor, Karim Douch, Nicolas Longépé, Zhen He, and Patrick Ebel. Off to new shores: A dataset & benchmark for (near-) coastal flood inundation forecasting. *Advances in Neural Information Processing Systems*, 37: 114797–114811, 2024.

[47] Qingsong Xu, Leon Frederik De Vos, Yilei Shi, Nils Rüther, Axel Bronstert, and Xiao Xiang Zhu. Urban flood modeling and forecasting with deep neural operator and transfer learning. *Journal of Hydrology*, page 133705, 2025.

[48] Qingsong Xu, Yilei Shi, Jie Zhao, and Xiao Xiang Zhu. Floodcastbench: A large-scale dataset and foundation models for flood modeling and forecasting. *Scientific Data*, 12 (1):431, 2025.

[49] Jiarui Yang, Kai Liu, Ming Wang, Gang Zhao, Wei Wu, and Qingrui Yue. A convolutional neural network-weighted cellular automaton model for the fast prediction of urban pluvial flooding processes. *International Journal of Disaster Risk Science*, 15(5): 754–768, 2024.

[50] PE Zope, TI Eldho, and V Jothiprakash. Impacts of urbanization on flooding of a coastal urban catchment: a case study of mumbai city, india. *Natural Hazards*, 75(1):887–908, 2015.

[51] PE Zope, TI Eldho, and Vinayakam Jothiprakash. Impacts of land use–land cover change and urbanization on flooding: A case study of oshiwara river basin in mumbai, india. *Catena*, 145:142–154, 2016.