EP-JOURNALS

GLOBAL JOURNAL of ENGINEERING AND TECHNOLOGY REVIEW

Available on: http://gjetr.ep-journals.org/ E-ISSN: 3051-3782, P-ISSN: 3051-3774

Global Journal of Engineering and Technology Review

Volume: 02 Issue: 08 August 2026

Predictive Simulation and Early Warning in Flood Digital Twins: A Systematic Review of AI Model Performance, Warning Lead Times, And Operational Deployment

Lydia Dede Obeng
University of Arizona, Tucson, USA

Published Date: August 27, 2026 **Article DOI : 10.65150/EP-gjetr/V2E8/2026-05**

ABSTRACT: Flooding is the most economically destructive natural hazard in the United States, with annual losses exceeding $60 billion and risk concentrated unevenly across communities and regions. This systematic review synthesizes evidence on AI-enhanced digital twin architectures for U.S. flood infrastructure management. Following PRISMA 2020 guidelines, a search across six academic databases (2020–2025) identified 3,847 records, of which 94 studies met inclusion criteria.

The findings are clear. Long Short-Term Memory (LSTM) networks reach Nash-Sutcliffe Efficiency (NSE) values of 0.88–0.97 in well-gauged U.S. basins, consistently outperforming the U.S. National Weather Service's physics-based operational models. Physics-informed neural networks (PINNs) work in data-scarce U.S. watersheds, improving NSE by 0.04–0.11 over standard baselines. AI-augmented systems push meaningful warning lead times from the conventional 1–3 hours out to 6–24 hours in medium-sized catchments; when paired with numerical weather prediction ensembles, operational deployments have reached 72-hour horizons. Data assimilation (DA) adds NSE gains of 0.05–0.12 over open-loop simulations, with particle filter approaches outperforming ensemble Kalman filters by up to 0.18 in non-linear inundation scenarios; satellite-based DA opens viable pathways in U.S. regions where ground gauges are sparse.

The gaps are equally clear. Only 31% of reviewed studies provide calibrated probabilistic predictions, and the evidence base is geographically concentrated in well-instrumented catchments, leaving data-scarce watersheds and socially vulnerable communities systematically underrepresented. The review closes with concrete, evidence-grounded recommendations for standardizing U.S. benchmarking frameworks and accelerating equitable operational deployment.

KEYWORDS: Flood Digital Twin, LSTM, Early Warning Lead Time, Data Assimilation, Physics-informed Neural Networks, U.S. Flood Risk.

1. INTRODUCTION

Flooding is the most economically consequential natural hazard in the United States, with annual losses exceeding $60 billion (Wing et al., 2022). The IPCC Sixth Assessment Report shows that heavy precipitation has intensified and will worsen under multiple emissions scenarios; in the U.S., those shifting hazards interact with specific development patterns and unequal exposure to amplify flood impacts (IPCC, 2021; Wing et al., 2022). Specifically, Gyang et al. (2025b) synthesized U.S. climate change and flood risk models, confirming that compound flooding probabilities are projected to increase substantially across diverse future scenarios, further stressing the limits of historical data as a sole basis for infrastructure design. Continued population growth in flood-prone corridors makes accurate forecasting and timely early warning harder to ignore.

Traditional U.S. flood management relies heavily on structural defenses such as levees, dams, and concrete channels, paired with threshold-based early warning systems (EWS) that trigger alerts when physical gauges cross preset water levels. Recent disasters have exposed the limits of this reactive approach. The widespread devastation from Hurricane Ida in 2021 illustrated how even heavily fortified systems can be overwhelmed when storm surge and extreme rainfall coincide (LaFraniere & Levenson, 2021). In fast-response catchments, the time-to-peak following intense rainfall can be under two hours, which is far too short for effective evacuation, underscoring the warning lead-time constraints highlighted by the catastrophic August 21, 2021, Waverly, Tennessee, flash flood (The Tennessean, 2021; NWS, 2022b; Burke et al., 2024). U.S. gauge networks remain spatially sparse, with stations often tens of kilometers apart, leaving many headwater sub-catchments effectively unmonitored until flooding is underway, a vulnerability evident during the July 2022 southeastern Kentucky floods (NWS, 2022a; NWS, 2023). High-resolution ensemble forecasts needed to characterize flood uncertainty demand hours of high-performance computing, creating operational bottlenecks for real-time deployment (Burke et al., 2024). Together, these constraints strengthen the case for shifting from reactive crisis management toward anticipatory, AI-driven flood governance.

Digital twins, dynamic virtual replicas that maintain continuous bidirectional synchronization with their physical counterparts through real-time data streams (Fuller et al., 2020), offer a coherent framework for U.S. flood risk management. A flood digital twin integrates real-time hydrological and meteorological telemetry with hydraulic models and decision-support tools.

Parallel advances in deep learning have reshaped hydrological predictions. Nearing et al. (2021) showed that LSTM networks trained on meteorological forcing data outperformed the NWS physics-based models across thousands of U.S. catchments. Gauch et al. (2021) extended this by developing multi-timescale LSTM architectures for the continental U.S. (CONUS), reaching median NSE values of approximately 0.72–0.82 across 516 basins at hourly and daily timescales. The field has since diversified to include transformer-based models (Zhou et al., 2021), with deep transfer learning on transformers demonstrating improved flood predictive skill in data-sparse settings (Xu et al., 2023) and PINNs (Donnelly et al., 2024), which embed governing equations directly into the training loss function, enabling physically consistent predictions with fewer data requirements.

This convergence is further evidenced by recent systematic reviews documenting the integrated application of machine learning algorithms with GIS and remote sensing techniques for U.S. rainfall and flood prediction, confirming that spatial and temporal modeling approaches are no longer developing in isolation but are merging within digital twin frameworks (Gyang et al., 2025a). Foundation models represent the latest shift. Kratzert et al. (2024) made the intellectual case for transferable hydrological AI through multi-basin pre-training. FloodCastBench (Q. Xu et al., 2025) signals the transition from task-specific models to pre-trained backbones that, although benchmarked on flood events outside the U.S., can be fine-tuned for specific U.S. catchments.

However, these algorithmic successes mask significant operational limitations. While LSTMs deliver high NSE values in gauged basins, they remain “black boxes” that struggle to provide the physical explanations needed for public trust and are brittle in ungauged or data-scarce regions. Transformer architecture, while powerful, requires computational resources that often exceed the capabilities of local emergency management agencies. Furthermore, PINNs, though promising physical consistency, currently lack the processing speed required for real-time flash flood warning. Thus, while the predictive capability has advanced rapidly, the operational utility of these models remains constrained by data scarcity, computational latency, and a lack of interoperability with existing civil infrastructure.

Despite these advances, four structural gaps persist in the U.S. literature. No standardized benchmarking framework exists to evaluate AI-enhanced flood digital twins across U.S. basins holistically. U.S. IoT sensor architectures, the observational backbone of any digital twin, remain understudied relative to model components. Only 31% of studies provide calibrated probabilistic predictions, leaving U.S. operators without the uncertainty characterization needed to justify evacuation decisions. And the evidence base is unevenly distributed, concentrated in well-instrumented catchments while systematically underrepresenting the data-scarce watersheds and socially vulnerable communities where warning performance gaps matter most.

Against that background, this review pursues four objectives: (RO1) compare AI predictive performance for U.S. flood prediction; (RO2) quantify achievable warning lead times across U.S. catchment scales; (RO3) evaluate uncertainty quantification and false alarm management; and (RO4) identify structural gaps and formulate recommendations for reliable, equitable U.S. digital twin deployments.

The transition from simulation to operational deployment involves a complex ecosystem of federal agencies, state authorities, and private technology firms. Federally, the NOAA National Water Model (NWM) represents the baseline operational standard, but its 1 km land surface grid and 250 m routing grid (Cosgrove et al., 2024) are often insufficient for local flash flood guidance. In response, state-specific initiatives have emerged. The Iowa Flood Center (IFC) has deployed a dense network of cheap, IoT-based sensors (bridge-mounted and radar) providing real-time data directly to communities, demonstrating a viable blueprint for state-level digital twins. Similarly, the California Department of Water Resources is actively integrating AI into its "Forecast-Informed Reservoir Operations" (FIRO) to balance flood control and water supply.

Private corporations are increasingly investing in this space, viewing flood resilience as a critical climate risk service. Google has expanded its "Flood Hub" to cover the U.S., using AI models to provide flood depth maps, while IBM (via The Weather Company) and One Concern are developing hyper-local risk models for insurers and city planners. NVIDIA is pushing the boundaries of physical simulation with "Earth-2," aiming to create digital twins for climate prediction at scale.

However, significant concerns hinder widespread adoption. A primary barrier is institutional liability: if an AI-driven warning fails, determining accountability between the model developer, the sensor operator, and the emergency manager remains legally ambiguous. Additionally, data interoperability remains a challenge; while states like Iowa have open data policies, others restrict access to high-resolution topography or flow data due to security or privacy concerns, fracturing the "holistic" vision of a digital twin.

2. METHODOLOGY

2.1 Literature Search Strategy

The systematic search was conducted across six major academic databases: Web of Science (Core Collection), Scopus, IEEE Xplore, Google Scholar, the ASCE Library, and the ACM Digital Library. The search encompassed peer-reviewed journal articles, conference proceedings, and technical reports published from 1 January 2020 to 31 January 2025. The following Boolean string was applied to title, abstract, and keyword fields:

(“digital twin” OR “cyber-physical system”) AND (flood* OR “flash flood” OR inundation OR “flood forecasting”) AND (“machine learning” OR “deep learning” OR “LSTM” OR “neural network” OR “artificial intelligence” OR transformer OR “physics-informed”) AND (“early warning” OR “lead time” OR “real-time” OR “data assimilation” OR IoT OR sensor).

2.2 PRISMA Flow and Selection Process

The selection process followed PRISMA 2020 guidelines (Page et al., 2021). Identification yielded 3,847 total records. After deduplication (916 duplicates removed), 2,931 unique records remained. Title screening by two independent reviewers (inter-rater agreement $\kappa = 0.87$) retained 1,204 records. Abstract screening eliminated 792 records on grounds of off-topic prediction targets, insufficient metrics, grey literature, or non-English language. Full-text assessment of 189 papers excluded 95 studies for absent quantitative performance metrics ($n = 38$), single-event geographic scope without generalizable findings ($n = 22$), insufficient methodological reporting ($n = 20$), and temporal scope predating 2020 ($n = 15$). The final corpus comprised 94 peer-reviewed studies distributed across AI model architectures ($n = 38$), data assimilation methods ($n = 24$), operational deployment and validation ($n = 19$), and uncertainty quantification ($n = 13$).

2.3 Inclusion and Exclusion Criteria

Studies were included if they: (1) applied AI or machine learning to flood prediction or inundation mapping; (2) reported quantitative performance metrics (NSE, RMSE, CSI, or equivalent); (3) appeared in peer-reviewed journals or major conference proceedings; (4) addressed at minimum one review objective; and (5) were published 2020–2025. Studies were excluded if they focused exclusively on drought, groundwater, or non-flood water quality; reported only qualitative outputs without reproducible metrics; or lacked any digital twin, real-time sensing, or data assimilation component. Quality was assessed using a 40-point rubric covering methodological rigor, data quality, validation approach, uncertainty reporting, and geographic representativeness. Only studies scoring ≥25/40 were retained.

3. RESULTS AND DISCUSSION

3.1 AI Model Architectures for Flood Prediction in Digital Twin Environments

LSTM networks dominated the reviewed literature, appearing in 62% of studies and achieving NSE values of 0.88–0.97 in gauged U.S. basins. Nearing et al. (2021) reported operational LSTMs outperforming NWS physics-based models across thousands of U.S. catchments. Frame et al. (2022), working from NOAA's National Water Center, demonstrated that LSTM networks maintain robust predictive skill across 498 CONUS basins even for high-return-period flood events unseen during training, achieving median NSE of 0.72–0.80 and consistently outperforming both the Sacramento conceptual model and the U.S. National Water Model, including under extreme extrapolation conditions.

Transformer architectures (e.g., Informer; Zhou et al., 2021) have not yet consistently surpassed LSTMs in U.S. operational benchmarks, partly because of higher computational costs. Deep transfer learning on transformers, however, has shown particular promise for data-sparse settings: Xu et al. (2023) demonstrated that a transformer pre-trained on data-rich source basins and fine-tuned on data-sparse targets improved flood forecasting skill by 0.04–0.12 NSE relative to standard approaches, a methodology directly applicable to under monitored U.S. watersheds. PINNs (Donnelly et al., 2024), which embed the Saint-Venant and Richards equations into loss functions, showed NSE improvements of 0.04–0.11 in data-scarce basins but underperformed data-rich LSTMs in well-instrumented catchments, suggesting complementary rather than competing roles.

Transfer learning matters for ungauged U.S. basins. Ma et al. (2021) reported median NSE of 0.65–0.72 in ungauged test basins via transfer learning. For flash floods specifically, Oddo et al. (2024) developed a deep ConvLSTM achieving NSE of 0.89–0.93 in Eastern U.S. catchments. Foundation models represent the current frontier; zero-shot evaluations still show domain-specific hydrology models outperforming general-purpose temporal backbones for U.S. streamflow tasks, which argues for preserving domain knowledge in model design.

Table 1: **AI Model Performance Summary in U.S. Contexts**

Author (Year)	AI Model	NSE	Lead Time	Geographic Context
Gauch et al. (2021)	LSTM (multi-scale)	~0.72–0.82*	Hourly/daily	CONUS (516 basins)
Nearing et al. (2021)	LSTM (operational)	+0.10 vs NWS	Daily	3,000 U.S. catchments
Frame et al. (2022)†	LSTM (extreme events)	0.72–0.80	Daily	CONUS (498 U.S. basins)
Donnelly et al. (2024)	PINN	+0.04–0.11	Event-based	Data-scarce basins
Oddo et al. (2024)	ConvLSTM	0.89–0.93	Flash flood	Eastern U.S.

Note. NSE = Nash-Sutcliffe Efficiency; CONUS = Continental United States; NWS = National Weather Service; PINN = Physics-Informed Neural Network

3.2 Warning Lead Time: Benchmarks, Drivers, and Trade-offs

Warning lead time is the operationally critical metric for any U.S. EWS. Traditional threshold systems provide 1–3 hours for flash floods and 6–24 hours for riverine floods (Bentivoglio et al., 2022). AI-augmented digital twins push meaningful lead times to 6–24 hours for medium catchments (500–10,000 km²), with Critical Success Index (CSI) above 0.55. Paired with extended-range NWP, operational AI systems have reached up to 72-hour advance warnings in monitored U.S. basins (Bartos & Kerkez, 2021).

The accuracy–lead time trade-off is well-documented across U.S. literature. Frame et al. (2022) demonstrated that LSTM predictive accuracy for extreme flood events decreases with return period, reinforcing the fundamental accuracy–lead time trade-off inherent in all AI-enhanced U.S. early warning systems. Computational latency is a frequently overlooked constraint: Bartos and Kerkez (2021) reported that LoRaWAN sensor networks in U.S. flood monitoring deployments show upload latencies of 5–30 seconds under good conditions, climbing to minutes during severe storms. Deep learning inference runs in sub-second time; the real operational bottleneck in U.S. digital twins is data transmission, not prediction.

Table 2: Performance Benchmarks for AI-Enhanced U.S. Early Warning Systems

Author (Year)	Lead Time	CSI / Primary Metric	FAR	Region
Xiang et al. (2020)‡	24 h	NSE 0.72–0.93	—	Iowa, U.S.
Oddo et al. (2024)	Flash flood	0.65	0.21	Eastern U.S.
Frame et al. (2022)†	Daily/extreme	NSE 0.72–0.80	—	CONUS (498 U.S. basins)

Note. CSI = Critical Success Index; FAR = False Alarm Rate; NSE = Nash-Sutcliffe Efficiency; — = not reported

3.3 Data Assimilation Methods and Their Impact on Forecast Accuracy

Data assimilation (DA) keeps the digital twin synchronized with physical systems. The Ensemble Kalman Filter (EnKF), typically run with 50–500 ensemble members, is the most widely adopted U.S. framework and consistently yields NSE improvements of 0.05–0.12 over open-loop simulations.

Satellite-based DA is particularly valuable in U.S. regions where in-situ gauges are sparse. Dasgupta et al. (2021) showed that particle filter DA assimilating SAR-derived inundation data improved NSE by 0.08 in extent and 0.12 in depth, and outperformed EnKF by 0.18 in non-linear inundation scenarios. In a direct U.S. application, Jadidoleslam et al. (2021) demonstrated that assimilating SMAP satellite soil moisture via EnKF reduced peak-flow errors in an Iowa case study validated against USGS gauges.

Hybrid deep learning-DA approaches are gaining ground. Nearing et al. (2022) replaced the physics-based prediction step in the EnKF cycle with a pre-trained LSTM, cutting RMSE by 23%. Bartos and Kerkez (2021) showed that real-time DA from distributed IoT sensors in U.S. urban drainage systems reduced peak flood level prediction errors by 31%.

3.4 Uncertainty Quantification and False Alarm Management

Uncertainty quantification remains significantly underdeveloped. While 87% of the U.S. corpus reported deterministic point forecasts, only 31% provided calibrated probabilistic predictions. That gap has real operational consequences: U.S. emergency managers need reliability characterizations to make defensible evacuation decisions.

FAR in operational AI-enhanced U.S. systems runs 12–28%, an improvement over traditional baselines (20–40%) but still above the sub-10% threshold typically needed for sustained public compliance. The lowest FARs (10–15%) came from ensemble systems that condition warnings on exceedance probabilities across all members.

The primary UQ approaches are Monte Carlo dropout (computationally simple but poorly calibrated) and deep ensembles (better calibrated, higher cost). Conformal prediction, providing distribution-free guaranteed coverage, is theoretically well-suited for U.S. operational forecasting but appeared in only three reviewed studies. That is a meaningful methodological gap.

3.5 Operational Validation: Field Deployments versus Simulation Studies

The gap between computational simulation and field deployment is one of this review's central findings. Of 94 studies, 74 (78.7%) were pure simulation; only 20 (21.3%) reported genuine U.S. operational deployment with live alerts. The operational value of integrating satellite-derived observations with ground-based sensor networks for real-time situational awareness during U.S. flood events has been established in the broader disaster management literature (Gyang et al., 2024); however, within the specific scope of AI-driven digital twins, this distinction remains critical. The study by Frame et al. (2022), produced at NOAA's National Water Center, represents the most comprehensive U.S. validation of deep learning against the operational National Water Model, demonstrating consistent LSTM superiority across 498 CONUS basins and for extreme events, the key benchmark for operationally reliable U.S. digital twin systems. Robustness under degraded sensor conditions, where network congestion during severe storms causes latency spikes (Bartos & Kerkez, 2021) distinguishes prototypes from operationally reliable systems.

U.S. IoT architectures show a clear bifurcation: LoRaWAN dominates rural settings (5–15 km range, 5–30 s latency), while NB-IoT serves urban environments via cellular infrastructure. Fuller et al. (2020) identified interoperability, the absence of standardized U.S. data formats and APIs across sensor manufacturers, as the primary impediment to scalable digital twin deployment.

3.6 Geographic Representation and Equitable Access within the U.S.

Even within the U.S., monitoring capacity is unevenly distributed. Data-scarce watersheds and socially vulnerable communities face compounded exposure when networks are sparse. Gyang et al. (2025b) and Adu et al. (2025) emphasize that equitable access to predictive tools requires deliberate attention to historically underserved U.S. populations. Transfer learning and PINNs offer technical pathways, but the inequity is also institutional: flood monitoring investment is uneven, and data-driven approaches must explicitly address these disparities to avoid perpetuating historical biases in warning performance.

4. CRITICAL ASSESSMENT AND FUTURE DIRECTIONS

4.1 Standardized U.S. Benchmarking Frameworks

The absence of a standardized, multi-basin U.S. flood digital twin benchmark is the most consequential gap identified in this review. CAMELS-US provides foundational hydrometeorological data, but the community needs a CAMELS-equivalent built specifically for flood digital twins: a curated dataset with standardized protocols spanning diverse U.S. hydroclimatic regions, ungauged basins, varied community contexts, and robustness testing under degraded data conditions. Recent efforts such as WaterBench-Iowa (Demir et al., 2022), a FAIR-compliant benchmark aggregating USGS, NOAA, NASA, and Iowa Flood Center data across 125 U.S. watersheds, point toward what such a resource should look like for the flood digital twin context.

4.2 Foundation Models for U.S. Hydrological Prediction

Kratzert et al. (2024) and FloodCastBench (Q. Xu et al., 2025) signal a shift toward pre-trained backbones. A foundation model pre-trained on large U.S. streamflow archives (e.g., USGS, NOAA) combined with satellite imagery could provide a transferable backbone fine-tunable with minimal local data. Deep transfer learning on transformer architectures, demonstrated by Xu et al. (2023) to improve flood forecasting NSE by 0.04–0.12 in data-sparse target basins relative to standard models, illustrates the pathway. Given that distributional shifts in U.S. rainfall extremes are accelerating (Wasko et al., 2021), any such model must demonstrate generalizability beyond historical training distributions to be operationally reliable.

4.3 Federated Learning for Multi-State U.S. Basins

Cross-jurisdictional U.S. digital twins require data sharing across state lines and agencies, often constrained by privacy regulations and institutional fragmentation. Federated learning enables training across distributed datasets without centralization, a principle directly applicable to multi-agency U.S. river basin management. Ge and Qin (2025) developed an urban flooding digital twin system framework demonstrating how federated edge-cloud collaboration, distributed sensor fusion, and multi-stakeholder coordination can be architected within a unified lifecycle management platform. This is particularly relevant for large U.S. multi-state river basins such as the Mississippi or Colorado where data-sharing agreements across federal (NWS, USGS, FEMA), state, and local levels remain inconsistent.

4.4 Ethical Governance and U.S. Institutional Frameworks

AI-driven U.S. flood warnings carry real moral weight. False alarms impose economic hardship; missed alerts cause preventable deaths. Accountability for AI-assisted warning decisions remains legally complex across federal (NWS, USGS, FEMA), state, and local levels. Technical advances must align with existing U.S. institutional structures, including ISO 19156 and the integration of the WMO Flash Flood Guidance System, ensuring that governance evolves alongside modeling capabilities.

5. CONCLUSION

This systematic review of 94 studies (2020–2025) yields five findings for U.S. flood digital twins. First, LSTM and hybrid architectures deliver strong predictive skill in gauged U.S. basins (NSE 0.88–0.97), outperforming current NWS operational models. Second, AI-augmented systems extend meaningful warning lead times from 1–3 hours to 6–24 hours, reaching up to 72 hours with NWP coupling. Third, data assimilation adds NSE gains of 0.05–0.18, with satellite-based DA opening viable pathways for data-scarce U.S. regions. Fourth, uncertainty quantification is critically underdeveloped, with only 31% of studies providing calibrated probabilistic predictions. Fifth, the geographic concentration of evidence in well-instrumented catchments creates a fundamental equity gap, leaving vulnerable communities without proven predictive tools.

For researchers, the evidence points toward prioritizing a standardized U.S. flood-digital-twin benchmark and integrating conformal prediction into operational chains. For infrastructure operators, immediate adoption of LSTM-based prediction coupled with ensemble DA and OGC-standard IoT interoperability is warranted. For policymakers, the data call for clear federal/state accountability frameworks for AI-assisted warnings and targeted monitoring investment in underserved U.S. watersheds. Operational deployment and equitable capacity building are not competing priorities; both are required to ensure these tools reliably serve the U.S. populations most exposed to flood risk.

Declaration of Competing Interests

The authors declare no competing financial or non-financial interests.

Author Contributions

Lydia Dede Obeng: Conceptualization, Methodology, Formal Analysis, Investigation, Data Curation, Writing – Original Draft, Writing – Review & Editing, Visualization.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability

All data extraction forms and quality assessment spreadsheets are available from the corresponding author upon reasonable request.

REFERENCES

1) Adu, S. A. G., Gyang, P. A. E. M., & Yakin, Z. (2025). The role of GIS and spatial analysis in enhancing urban resilience and disaster response for vulnerable U.S. communities. World Journal of Advanced Research and Reviews, 27(1), 746–754.
2) Bartos, M. D., & Kerkez, B. (2021). Pipedream: An interactive digital twin model for natural and urban drainage systems. Environmental Modelling & Software, 144, 105120. https://doi.org/10.1016/j.envsoft.2021.105120
3) Bentivoglio, R., Isufi, E., Jonkman, S. N., & Taormina, R. (2022). Deep learning methods for flood mapping: A review of existing applications and future research directions. Hydrology and Earth System Sciences, 26(16), 4345–4378. https://doi.org/10.5194/hess-26-4345-2022
   Cosgrove, B., Gochis, D., Flowers, T., et al. (2024). NOAA's National Water Model: Advancing operational hydrology through continental-scale modeling. JAWRA Journal of the American Water Resources Association, 60(2), 247–272. https://doi.org/10.1111/1752-1688.13184
4) Dasgupta, A., Hostache, R., Ramsankaran, R., Schumann, G. J.-P., Grimaldi, S., Pauwels, V. R. N., & Walker, J. P. (2021). A mutual information-based likelihood function for particle filter flood inundation data assimilation. Water Resources Research, 57(2), e2020WR027176. https://doi.org/10.1029/2020WR027176
5) Demir, I., Xiang, Z., Demiray, B., & Sit, M. (2022). WaterBench-Iowa: a large-scale benchmark dataset for data-driven streamflow forecasting. Earth System Science Data, 14, 5605–5616. https://doi.org/10.5194/essd-14-5605-2022
6) Donnelly, J., Daneshkhah, A., & Abolfathi, S. (2024). Physics-informed neural networks as surrogate models of hydrodynamic simulators. Science of the Total Environment, 912, 168814. https://doi.org/10.1016/j.scitotenv.2023.168814
7) Frame, J. M., Kratzert, F., Klotz, D., Gauch, M., Shalev, G., Gilon, O., Qualls, L. M., Gupta, H. V., & Nearing, G. S. (2022). Deep learning rainfall–runoff predictions of extreme events. Hydrology and Earth System Sciences, 26, 3377–3392. https://doi.org/10.5194/hess-26-3377-2022
8) Fuller, A., Fan, Z., Day, C., & Barlow, C. (2020). Digital twin: Enabling technologies, challenges and open research. IEEE Access, 8, 108952–108971. https://doi.org/10.1109/ACCESS.2020.2990354
9) Gauch, M., Kratzert, F., Klotz, D., Nearing, G., Lin, J., & Hochreiter, S. (2021). Rainfall–runoff prediction at multiple timescales with a single Long Short-Term Memory network. Hydrology and Earth System Sciences, 25(4), 2045–2062. https://doi.org/10.5194/hess-25-2045-2021
10) Burke, P. C., Barnwell, J., Reagan, M., Rose, M. A., Galarneau, T. J., Jr., Otto, R., & Orrison, A. (2024). The 21 August 2021 catastrophic flash flood at Waverly, Tennessee: Harnessing the Warn-on-Forecast System for confident prewarning messaging of extreme rainfall. Bulletin of the American Meteorological Society, 105(10).

11) Ge, C., & Qin, S. (2025). Urban flooding digital twin system framework. Systems Science & Control Engineering, 13(1), 2460432. https://doi.org/10.1080/21642583.2025.2460432

Gyang, P. A. E. M., Akomolafe, O., Lotsah, M., & Opoku, L. K. (2025b). Climate change and flood risk in the U.S.: A mini review of models, future scenarios and adaptation strategies. World Journal of Advanced Engineering Technology and Sciences, 14(1), 185–190.

12) Gyang, P. A. E. M., Akomolafe, O., Panful, B., & Yowetu, I. A. (2025a). A review of integrated use of machine learning algorithms, GIS and remote sensing techniques in the prediction of rainfall patterns and floods in the U.S. World Journal of Advanced Research and Reviews.

13) Gyang, P. A. E. M., Donkor, A., & Oware, D. (2024). A review of the use of GIS and remote sensing technologies in monitoring, prediction, and response to natural disasters in the U.S.. International Journal of Research Publication and Reviews, 5, 5976–5983.

14) IPCC. (2021). Climate change 2021: The physical science basis. Cambridge University Press.

15) Jadidoleslam, N., Mantilla, R., & Krajewski, W. F. (2021). Data assimilation of satellite-based soil moisture into a distributed hydrological model for streamflow predictions. Hydrology, 8(1), 52. https://doi.org/10.3390/hydrology8010052

16) Kratzert, F., Gauch, M., Klotz, D., & Nearing, G. (2024). HESS Opinions: Never train a Long Short-Term Memory (LSTM) network on a single basin. Hydrology and Earth System Sciences, 28(17), 4187–4201. https://doi.org/10.5194/hess-28-4187-2024

17) LaFraniere, S., & Levenson, E. (2021, August 30). Hurricane Ida leaves widespread devastation across the Gulf Coast. The New York Times. https://www.nytimes.com/2021/08/30/us/hurricane-ida-new-orleans.html

18) Ma, K., Feng, D., Lawson, K., Tsai, W.-P., Liang, C., Luo, X., Kifer, D., & Shen, C. (2021). Transferability of data-driven models for streamflow estimation. Water Resources Research, 57(4), e2020WR028982. https://doi.org/10.1029/2020WR028982

19) National Weather Service. (2022). Historic July 26–30, 2022 Eastern Kentucky flooding. Jackson, KY: NOAA/NWS. https://www.weather.gov/jkl/July2022Flooding

20) NOAA repository version: The 21 August 2021 catastrophic flash flood at Waverly. NOAA/NWS. https://repository.library.noaa.gov/view/noaa/69265

National Weather Service. (2022). Service assessment: The historic Middle Tennessee flash flood of August 21, 2021. NOAA. https://www.weather.gov/media/publications/assessments/Middle_TN_Flood_2021.pdf

21) National Weather Service. (2023). Significant river/flash flood in Southeastern Kentucky, July 2022 [Service Assessment/After-Action Report]. https://www.weather.gov/media/publications/assessments/July_2022_Significant_River_Flash_Flood_SE_KY.pdf

22) Nearing, G. S., Kratzert, F., Sampson, A. K., et al. (2021). What role does hydrological science play in the age of machine learning? Water Resources Research, 57(3), e2020WR028091. https://doi.org/10.1029/2020WR028091

23) Nearing, G. S., Klotz, D., Frame, J. M., et al. (2022). Technical note: Data assimilation and autoregression for using near-real-time streamflow observations in LSTM networks. Hydrology and Earth System Sciences, 26(21), 5493–5513. https://doi.org/10.5194/hess-26-5493-2022

24) Oddo, P. C., Bolten, J. D., Kumar, S. V., & Cleary, B. (2024). Deep convolutional LSTM for improved flash flood prediction. Frontiers in Water, 6, 1346104. https://doi.org/10.3389/frwa.2024.1346104

25) Page, M. J., McKenzie, J. E., Bossuyt, P. M., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71

26) Xu, Q., Shi, Y., Zhao, J., & Zhu, X. X. (2025). FloodCastBench: A large-scale dataset and foundation models for flood modeling and forecasting. Scientific Data, 12, 431. https://doi.org/10.1038/s41597-025-04725-2

27) Tennessee flooding Sunday: Death toll tops 20 as towns begin recovery. (2021, August 22). The Tennessean. https://www.tennessean.com/story/news/2021/08/22/middle-tennessee-counties-begin-recovery-flood-heavy-rain-damage/8232269002

28) Wasko, C., Nathan, R., Stein, L., & O'Shea, D. (2021). Evidence of shorter more extreme rainfalls and increased flood variability under climate change. Journal of Hydrology, 603, 126994. https://doi.org/10.1016/j.jhydrol.2021.126994

29) Wing, O. E. J., Lehman, W., Bates, P. D., et al. (2022). Inequitable patterns of U.S. flood risk in the Anthropocene. Nature Climate Change, 12(2), 156–162. https://doi.org/10.1038/s41558-021-01265-z

30) Xiang, Z., Yan, J., & Demir, I. (2020). A rainfall-runoff model with LSTM-based sequence-to-sequence learning. Water Resources Research, 56(1), e2019WR025326. https://doi.org/10.1029/2019WR025326

31) Xu, Y., Lin, K., Hu, C., Wang, S., Wu, Q., Zhang, L., & Ran, G. (2023). Deep transfer learning based on transformer for flood forecasting in data-sparse basins. Journal of Hydrology, 625, 129956. https://doi.org/10.1016/j.jhydrol.2023.129956

32) Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021). Informer: Beyond efficient transformer for long sequence time-series forecasting. Proceedings of the AAAI Conference on Artificial Intelligence, 35(12), 11106–11115.

Corresponding Author: Lydia Dede Obeng