



Machine Learning and Deep Learning Approaches for Flood Prediction

Chetan Pathe¹, Yash Borkar², Purva Waghmare³, Rishita Dudhe⁴, Vaishali Chaudhari⁵, Sushama Telrandhe⁶

^{1,2,3,4}UG Student Department of Artificial Intelligence, Guru Nanak Institute of Engineering and Technology
Nagpur, India

⁵Assistant Professor, Department of Artificial Intelligence Guru Nanak Institute of Engineering and Technology
Nagpur, India

⁶Associate Professor, Department of Artificial Intelligence Guru Nanak Institute of Engineering and Technology
Nagpur, India

Abstract

How to Cite this Article:

Pathe, C., Borkar, Y., Waghmare, P., Dudhe, R. & Chaudhari, V. (2026). Machine Learning and Deep Learning Approaches for Flood Prediction. International Journal of Creative and Open Research in Engineering and Management, 02(10), 1-9.
<https://doi.org/10.55041/ijcope.v2i10.030>

License:

This article is published under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

© The Author(s). Published by International Journal of Creative and Open Research in Engineering and Management.



<https://doi.org/10.55041/ijcope.v2i10.030>

Flooding remains one of the most damaging natural hazards worldwide, and the growing availability of hydrological sensor networks and satellite data has pushed researchers toward data-driven alternatives to conventional physics-based hydrological models. This paper synthesizes findings from six recent bodies of work — three literature reviews and three influential primary studies — covering the evolution of machine learning (ML) applications in flood and flash-flood forecasting, from early architecture benchmarking through real-world operational deployment to static susceptibility mapping. The synthesis traces the field's progression from early artificial neural network (ANN) and support vector machine (SVM) implementations toward the current dominance of Long Short-Term Memory (LSTM) networks and hybrid architectures that combine convolutional, recurrent, and physics-based components. Across the reviewed literature, meteorological and hydrological variables — principally rainfall and water level — account for the substantial majority of model inputs, while error-based metrics such as RMSE and Nash-Sutcliffe Efficiency (NSE) remain the standard for performance evaluation. The synthesis also identifies recurring structural weaknesses in the field: a heavy geographic concentration of research in East and Southeast Asia, limited dataset sharing that constrains reproducibility, underused approaches such as graph neural networks and transfer learning, and insufficient attention to

uncertainty quantification. The findings suggest that while data-driven flood forecasting has matured considerably in predictive capability, its transition into dependable operational early-warning systems still depends on resolving these gaps in transparency, generalizability, and interpretability.

Keywords— Flood forecasting; machine learning; deep learning; LSTM; hydrological modeling; hybrid models; flash floods.



INTRODUCTION

Flood events account for a disproportionate share of global disaster losses, and their frequency is widely expected to rise as precipitation patterns shift under climate change. For decades, flood prediction relied on physically based hydrological and hydraulic models that simulate rainfall-runoff processes according to governing physical equations. While these models remain scientifically well-grounded, they typically demand extensive calibration, dense monitoring networks, and considerable computational resources — conditions rarely met in ungauged or data-scarce basins, and particularly limiting for the short lead times required during fast-onset flash floods.

This gap between physical realism and operational responsiveness has motivated a sustained shift toward data-driven modeling. Rather than encoding the physics of a catchment, machine learning (ML) and deep learning (DL) techniques learn statistical relationships directly from historical rainfall, discharge, and water-level records, offering faster training, lower computational overhead, and the ability to capture nonlinear dependencies that classical statistical methods struggle to represent. This shift has not been a single event but a multi-decade evolution, moving from early neural network implementations in the 1990s, through the widespread adoption of support vector machines and ensemble tree methods in the 2000s and 2010s, to the current era dominated by recurrent architectures and hybrid deep learning pipelines.

This paper draws together three recent surveys of this literature [1]–[3] — a foundational review of ML methods for general flood prediction, a systematic review focused specifically on flash-flood forecasting, and a broader systematic mapping study of data-driven flood forecasting published between 2019 and 2024 — together with three widely cited primary studies that illustrate the trajectory the reviews describe: the original

benchmarking of LSTM networks against a conceptual hydrological model [4], Google's operational LSTM-based flood-forecasting framework deployed across South Asia [5], and an earlier GIS-based ensemble susceptibility-mapping study [6] that represents the pre-deep-learning, spatial-analysis branch of the field. Together, these six sources are used to characterize where the field currently stands, which technical approaches have proven most durable, and where meaningful gaps continue to limit real-world deployment.

LITERATURE REVIEW

Early data-driven flood modeling built on statistical time-series methods such as autoregressive moving average (ARMA), autoregressive integrated moving average (ARIMA), and multiple linear regression. These approaches offered computational simplicity but performed poorly against the nonlinear, threshold-driven dynamics that characterize rainfall-runoff behavior, motivating the adoption of ML techniques capable of learning such relationships directly from data.

The foundational review examined in this synthesis [1] surveyed more than 180 comparative studies published up to 2018 and identified artificial neural networks as the most consistently applied technique, alongside adaptive neuro-fuzzy inference systems (ANFIS), wavelet neural networks, support vector machines, and decision-tree variants including random forest and classification and regression trees. A central finding of that review was the steady rise of hybrid methods — combinations of ML algorithms with wavelet decomposition, ensemble averaging, or optimization routines — which by 2017 had come to outnumber single-method implementations in the literature by a wide margin, driven by their capacity to address the generalization weaknesses of standalone models.



The second body of work, a systematic review targeting flash floods specifically [2], examined studies published through January 2024 and found a much narrower technical picture: LSTM networks alone accounted for 60% of the reviewed implementations, with multilayer perceptrons a distant second. This concentration reflects the particular demands of flash-flood forecasting, where short lead times and rapid catchment response favor architectures explicitly designed to model sequential dependencies. That review also surfaced a research landscape skewed heavily toward China and, to a lesser extent, the Republic of Korea and the United States — countries with both severe flood exposure and established monitoring infrastructure.

The third and broadest study, a systematic mapping of 363 data-driven flood forecasting papers from 2019 to 2024 [3], largely reinforced these patterns while adding methodological texture. It confirmed LSTM's dominance (21% of all algorithm implementations) but also documented a marked rise in hybrid and ensemble approaches, which grew from a negligible share of publications in 2019 to roughly ten percent by 2024. This study further clarified several conceptual distinctions — forecast horizon, lead time, lag time, and look-back window — that are frequently used loosely across the literature but materially affect how models are designed and compared.

Across all three reviews, a consistent thread emerges: rainfall and water-level or discharge measurements form the backbone of nearly every model, geographic and remote-sensing variables are used mainly as supplementary inputs in hybrid configurations, and no single algorithm consistently outperforms all others across different basins, climates, and forecast horizons.

Three primary studies illustrate these patterns at the level of individual implementations rather than aggregate review statistics. Kratzert et al. [4] introduced the LSTM approach to rainfall-runoff modeling that the later reviews describe as dominant, training a single network across 241 CAMELS catchments and showing it could match or exceed the well-established Sacramento Soil

Moisture Accounting model coupled with the Snow-17 snow routine, while also demonstrating that a regionally trained LSTM could transfer learned process understanding to individual, data-scarce catchments. Nevo et al. [5] describe Google's operational flood-forecasting system, which has been running since 2018 and, during the 2021 monsoon season, provided real-time riverine flood warnings across large gauged rivers in India and Bangladesh; the system pairs LSTM-based stage forecasting with a novel machine-learning "manifold" model for inundation extent and depth, offering a concrete example of the transition from benchmark accuracy to deployed early-warning infrastructure that the reviewed literature identifies as still incomplete. Tehrany et al. [6], by contrast, represent the GIS-based susceptibility-mapping tradition that predates the LSTM era: rather than forecasting discharge over time, their ensemble of weights-of-evidence and support vector machine models combines static conditioning factors — elevation, slope, and distance from the river network among them — to classify which locations are most flood-prone, illustrating the interpretability-oriented, spatial-analysis line of work that the reviews note has been overshadowed by sequence models in recent years.

METHODOLOGY

This paper adopts a comparative desk-based synthesis methodology rather than primary data collection. Three peer-reviewed literature reviews were selected as the principal source material: a foundational review covering more than 180 comparative studies published up to 2018 [1], a systematic review targeting flash-flood forecasting specifically through January 2024 [2], and a systematic mapping study of 363 data-driven flood forecasting papers published between 2019 and 2024 [3]. These three sources were chosen because together they span the full technical evolution of the field — from early single-method ML approaches to the current dominance of LSTM-based and hybrid architectures — while also differing in scope, from general flood prediction to the more time-critical case of flash floods. Three additional primary studies [4]–[6] were incorporated to ground the reviews' aggregate



statistics in concrete, individually verifiable implementations: a foundational LSTM rainfall-runoff benchmarking study [4], a deployed operational forecasting system [5], and a GIS-based susceptibility-mapping study [6] representing an approach the reviews classify separately from time-series forecasting.

For each source review and each supporting primary study, the reported findings on model architecture prevalence, input variable categories, evaluation metrics, geographic distribution of research, and identified research gaps were extracted and tabulated for cross-comparison. Recurring patterns that appeared consistently across all three reviews — such as the dominance of rainfall and water-level inputs, the reliance on RMSE and Nash-Sutcliffe Efficiency for evaluation, and the geographic concentration of research in East and Southeast Asia — were treated as the most robust findings of the synthesis, while patterns unique to a single review were reported with appropriate caveats regarding their scope.

RESULTS AND DISCUSSION

The technical landscape described across the three reviews can be organized into four broad modeling families, each with distinct strengths and constraints. Single ML methods form the historical baseline of the field. Artificial neural networks, and their backpropagation and multilayer perceptron variants, were favored for their fault tolerance and ability to model rainfall-flood relationships without prior knowledge of catchment physics, though they were repeatedly noted to struggle with generalization beyond their training conditions. Support vector machines and their regression counterpart offered stronger performance for shorter lead times and smaller datasets, owing to their basis in structural risk minimization, while decision-tree-based methods such as random forest provided useful interpretability through feature-importance rankings, particularly for susceptibility mapping tasks rather than continuous time-series forecasting — an approach exemplified by the GIS-based ensemble of weights-of-evidence and support vector machine models developed by Tehrany et al. [6], which classified flood-prone

terrain from static conditioning factors rather than forecasting discharge over time.

Deep learning architectures, and LSTM networks above all, have become the field's default choice because their gated memory cells are explicitly suited to capturing the delayed, sequential dependencies between rainfall and streamflow response, a capability first demonstrated at scale by Kratzert et al. [4], whose regionally trained LSTM matched or exceeded a well-established conceptual hydrological model across more than 200 catchments. Convolutional neural networks contribute a complementary capability, extracting spatial features from gridded precipitation, radar, or terrain data, but are rarely used in isolation for time-series forecasting given their design orientation toward spatial rather than temporal patterns.

Hybrid and ensemble approaches represent the most active area of recent growth. These include architectures that pair CNNs with LSTM or GRU layers to capture spatiotemporal dynamics jointly, wavelet-decomposition front-ends that improve data quality before model training, and ensembles that aggregate multiple model outputs through bagging, boosting, or stacking to reduce variance and quantify uncertainty. A second hybrid category integrates data-driven components directly with physics-based models such as HEC-HMS, MIKE, or the Xinanjiang model, combining the interpretability and physical grounding of process-based simulation with the pattern-recognition strength of ML.

Beyond individual model architectures, the transition from benchmark accuracy to operational reliability is best illustrated by Nevo et al. [5], whose system layers LSTM-based stage forecasting, a linear baseline model, and two inundation models — a simpler thresholding approach and a novel "manifold" model capable of estimating both inundation extent and depth — into a four-subsystem pipeline covering data validation, forecasting, inundation modeling, and alert distribution. During the 2021 monsoon season this pipeline issued real-time warnings across large gauged rivers in India and Bangladesh, offering



one of the few documented cases in the reviewed literature where LSTM-based forecasting has been sustained in production rather than evaluated only on historical benchmark data.

Table I summarizes the key quantitative patterns reported across the six synthesized studies.

Table I: Key Quantitative Findings Across the Reviewed Studies

Metric	Reported Finding
LSTM share of flash-flood implementations	60% [2]
LSTM share of all algorithm instances (2019–2024)	21% [3]
Meteorological + hydrological share of inputs	~75% of variables [3]
Hybrid/ensemble share of studies, 2019 vs. 2024	~2% → ~10% [3]
China's share of flash-flood / mapping studies	~33%–35% [2], [3]
Studies with publicly available datasets	10% [2]
Flash-flood articles, 2000–2020 vs. 2021–2023	9 vs. 41 [2]
LSTM vs. conceptual model (CAMELS benchmark)	Matched/exceeded SAC-SMA+Snow-17 across 241 catchments [4]
Operational deployment coverage, 2021 monsoon	Real-time warnings across gauged rivers in India and Bangladesh [5]
Susceptibility-mapping model type	GIS ensemble of weights-of-evidence + SVM [6]

Data inputs across the reviewed studies fall into a small number of recurring categories: meteorological variables (chiefly rainfall and precipitation), hydrological variables (water level, discharge, and streamflow), and, less frequently, geographic and terrain data, land cover, soil properties, and remote-sensing imagery from

satellite or radar platforms. Despite the demonstrated value of remote sensing for capturing rainfall dynamics in ungauged basins, its actual use across the literature remains modest, a pattern attributed to coarse spatial resolution, sensor availability constraints, and the mismatch between remote-sensing update frequency and the high temporal resolution flash-flood forecasting requires.

Evaluation practice is comparatively uniform. Root Mean Squared Error and the Nash-Sutcliffe Efficiency coefficient dominate as reported metrics, followed by the coefficient of determination and mean absolute error. This concentration on aggregate error measures, while useful for general comparison, has been flagged as a limitation in its own right: such metrics are less sensitive to the accurate timing and magnitude of flood peaks, which are precisely the events an operational forecasting system exists to anticipate. Probabilistic and uncertainty-aware metrics remain comparatively rare across the surveyed studies.

Taken together, the three reviews describe a field expanding rapidly in both volume and technical sophistication. The earliest review, drawing on more than six thousand initial search results narrowed to 180 comparative studies, documented a clear multi-year trend in which hybrid ML methods overtook single-method approaches, attributing the shift to hybridization, data decomposition, ensemble averaging, and model optimization as the four most effective strategies for improving prediction quality. The flash-flood-specific review found an even sharper acceleration, with publication counts rising from just nine articles across the two decades from 2000 to 2020 to forty-one articles in the three years from 2021 to 2023 alone. The broadest mapping study corroborated this trajectory at a larger scale, recording a threefold increase in publications between 2019 and 2024, with hybrid and ensemble methods growing from roughly two percent of studies in 2019 to ten percent by 2024.

The geographic distribution of research effort was strikingly uneven across all reviews that examined it. China alone accounted for roughly a third to



over a third of studies in both the flash-flood review and the broader mapping study, with East and Southeast Asia collectively representing more than half of all research output. North America, Europe, and especially South America, Africa, and the Middle East were comparatively underrepresented, a pattern the reviews attribute to a combination of monitoring infrastructure limitations, funding disparities, and differing institutional approaches to flood management in economically advanced but less-represented regions.

Data transparency emerged as a shared concern. The flash-flood-specific review found that only 10% of the studies it examined made their datasets publicly available, a limitation that directly constrains reproducibility and fair comparison between competing methods. This lack of shared benchmarks was cited by all three reviews as a structural barrier to determining which architectures genuinely perform best under comparable conditions, since no single method was found to consistently outperform others once regional hydrology, data quality, and forecast horizon were accounted for.

CONCLUSION

The synthesized literature demonstrates that data-driven flood forecasting has moved well past a proof-of-concept stage, with LSTM-based and hybrid architectures now capable of producing forecasts that meet or exceed the accuracy of traditional physically based models across a wide range of lead times and hydrological settings. This progress has been driven less by any single algorithmic breakthrough than by a sustained pattern of hybridization — combining spatial and temporal feature extraction, decomposing noisy input signals, and blending physics-based structure with data-driven flexibility. The primary studies reviewed alongside the three surveys make this progression concrete: the field's move from single-basin LSTM benchmarking [4] to sustained operational deployment [5] spans barely half a decade, even as GIS-based ensemble methods for static susceptibility mapping [6] remain a parallel,

still-active line of work for hazard assessment rather than time-series prediction.

At the same time, the reviewed studies converge on a consistent set of unresolved issues. Reproducibility is undermined by the scarcity of shared datasets and standardized benchmarks, making genuine cross-study comparison difficult. Model interpretability remains limited, as the deep architectures that now dominate the field function largely as black boxes, offering little insight into the physical processes driving a given forecast — a concern with direct consequences for decision-makers who must justify evacuation or resource-allocation decisions. Geographic imbalance in research effort leaves many flood-prone but under-resourced regions without locally validated forecasting tools, and techniques such as graph neural networks, capable of modeling the spatial connectivity of river networks, along with transfer learning approaches suited to data-scarce basins, remain conspicuously underexplored relative to their apparent potential. Uncertainty quantification, essential for operational early-warning systems that must communicate confidence alongside a forecast, is similarly underdeveloped.

Future work identified across the reviewed literature points toward several converging priorities: the development of open benchmark datasets and standardized evaluation protocols, wider integration of physics-informed neural network approaches that constrain data-driven models with known hydrological relationships, greater use of explainable AI techniques to make model reasoning more transparent to hydrologists and emergency responders, and closer integration of forecasting models into operational early-warning infrastructure through sustained collaboration between ML researchers, hydrologists, and civil defense authorities. Addressing these gaps will likely determine whether the substantial predictive gains documented in the literature translate into measurably safer outcomes for flood-prone communities.



REFERENCES

- [1] A. Mosavi, P. Ozturk, and K. Chau, “Flood Prediction Using Machine Learning Models: Literature Review,” *Water*, vol. 10, no. 11, p. 1536, Oct. 2018.
- [2] L. B. L. Santos, E. V. Escobar-Silva, L. F. Satolo, R. S. Oyarzabal, M. M. Diniz, R. G. Negri, G. R. T. Lima, S. Stephany, J. A. J. P. Soares, J. S. Duque, F. L. Saraiva Filho, A. N. Caseri, and L. Bacelar, “Machine learning-based hydrological models for flash floods: a systematic literature review,” *Smart Construction and Sustainable Cities*, vol. 3, no. 21, Oct. 2025.
- [3] B. Kuhaneswaran, G. Sorwar, A. R. Alaei, and F. Tong, “Evolution of Data-Driven Flood Forecasting: Trends, Technologies, and Gaps—A Systematic Mapping Study,” *Water*, vol. 17, no. 15, p. 2281, Jul. 2025.
- [4] F. Kratzert, D. Klotz, C. Brenner, K. Schulz, and M. Herrnegger, “Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks,” *Hydrology and Earth System Sciences*, vol. 22, pp. 6005–6022, 2018.
- [5] S. Nevo, E. Morin, A. Gerzi Rosenthal, A. Metzger, C. Barshai, D. Weitzner, D. Voloshin, F. Kratzert, G. Elidan, G. Dror, et al., “Flood forecasting with machine learning models in an operational framework,” *Hydrology and Earth System Sciences*, vol. 26, no. 15, pp. 4013–4032, 2022.
- [6] M. S. Tehrany, B. Pradhan, and M. N. Jebur, “Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS,” *Journal of Hydrology*, vol. 512, pp. 332–343, 2014.