AI-Driven Predictive Modeling for Sustainable Urban Water Resource Management and Flood Risk Assessment

Lavkush Gupta, Nishikant Kumar,
Raushan Anand, Rajnish Kumar,
Raushan Kumar

Abstract: *Rapid urbanization, aging distribution infrastructure, and increasingly erratic rainfall patterns have placed municipal water systems under mounting strain, exposing cities to the simultaneous risks of supply shortage and pluvial flooding. This paper presents an integrated artificial-intelligence framework that couples short-term water demand forecasting with spatial flood risk classification within a single analytical pipeline. Historical consumption records, hydrometeorological variables, and land-cover indicators were combined and processed through an ensemble of Long Short-Term Memory networks, Random Forest regressors, and Extreme Gradient Boosting classifiers. The stacked ensemble achieved a coefficient of determination of 0.93 for demand prediction and a classification accuracy of 91.4 percent for flood risk zoning, outperforming individually tuned baseline models across every evaluation metric. Results indicate that antecedent rainfall, river gauge elevation, and impervious surface coverage are the dominant predictors of urban flood susceptibility, while historical demand lag and seasonal temperature govern consumption variability. The proposed framework offers municipal planners a computationally efficient, data-driven instrument for anticipating both water scarcity and inundation events, thereby supporting proactive infrastructure investment and emergency preparedness under climate uncertainty.*

Keywords: *Artificial Intelligence, Water Demand Forecasting, Flood Risk Assessment, Random Forest Regressors, Extreme Gradient Boosting Classifiers, Ensemble Learning.*

Lavkush Gupta*(guptalavkush1990@gmail.com), Assistant Professor, Department of civil Engineering, Gaya College of Engineering, Gaya, Bihar, 823003, India
Nishikant Kumar (nishikantk14171@gmail.com), Assistant Professor, Department of civil Engineering, B P Mandal College of Engineering Madhepura Bihar, 852128, India
Raushan Anand (raushan.anand8@gmail.com), Assistant Professor, Department of civil Engineering, B. P. Mandal College of Engineering, Madhepura, Bihar , 852128, India
Rajnish Kumar (profrajnishgecmunger@gmail.com), Assistant Professor, Department of civil Engineering, Government Engineering College Munger Bihar , 811202, India
Raushan Kumar (raushanmit2k8@gmail.com) , Assistant Professor, Department of civil Engineering, Gaya College of Engineering, Gaya, Bihar, 823003, India

1. Introduction

Cities absorb a disproportionate share of the pressures created by climate variability and demographic expansion. The United Nations World Water Development Report projects that the global urban population confronting water scarcity will roughly double, rising from 930 million residents in 2016 to somewhere between 1.7 and 2.4 billion by mid-century (UNESCO/UN-Water, 2023). At the same time, the paving-over of natural catchments and the densification of drainage networks have made metropolitan areas markedly more susceptible to flash flooding, even under moderate rainfall intensities. Municipal utilities therefore face a dual mandate: they must deliver a continuous, adequate water supply while simultaneously safeguarding residents from inundation, often with the same finite budget for sensors, personnel, and infrastructure upgrades.

Conventional planning tools have generally treated these two problems in isolation. Water demand has historically been projected using statistical time-series techniques such as autoregressive integrated moving average models, while flood hazard has been estimated through physically based hydrodynamic solvers that simulate rainfall-runoff transformation at coarse temporal resolution. Both approaches carry recognized limitations. Statistical demand models struggle to capture the nonlinear interactions between temperature, tariff structure, and consumer behaviour (Donkor et al., 2014), whereas hydrodynamic flood simulators, though physically grounded, are computationally expensive and difficult to update in near real time as urban land cover changes (Bermúdez et al., 2018).

Machine learning has begun to close this gap. Ensemble tree methods and recurrent neural architectures now match or exceed the accuracy of classical hydrological solvers while running orders of magnitude faster, a property that is particularly valuable for early-warning applications where decisions must be issued within minutes of a rainfall alert (Berkhahn et al., 2019). Extreme learning machines and long short-term memory networks have likewise demonstrated superior performance over linear regression and shallow neural networks in short-horizon water consumption forecasting (Mouatadid & Adamowski, 2017; Pu et al., 2023). Despite these individual advances, relatively few studies have attempted to unify demand forecasting and flood classification into a single operational workflow that shares data infrastructure, feature engineering routines, and evaluation protocols — an omission this study seeks to address.

This paper makes three contributions. First, it proposes a shared data-processing and feature-engineering pipeline that feeds two distinct predictive tasks, reducing redundancy in data governance for utilities that must otherwise maintain separate systems. Second, it benchmarks a stacked ensemble against individually optimized Random Forest, XGBoost, and LSTM baselines across both tasks, reporting standardized error and classification metrics. Third, it translates model outputs into a spatial risk map and an interpretable feature-importance ranking that non-specialist planners can use directly for zoning and capital-investment decisions.

2. Related Work

Research on data-driven urban water demand forecasting has evolved considerably over the past decade. Donkor et al. (2014) conducted a comprehensive review of forecasting methods and concluded that model choice should be guided by the forecast horizon, noting that autoregressive and regression-based techniques remain competitive for long-term planning but lose accuracy at hourly and daily resolution. Building on this observation, Mouatadid and Adamowski (2017) compared artificial neural networks, support vector regression, and extreme learning machines for the city of Montreal and found that extreme learning machines produced the lowest root-mean-square error at one- and three-day lead times. More recently, hybrid architectures that combine convolutional feature extraction with recurrent temporal modeling have further improved short-term accuracy; Pu et al. (2023) reported that a wavelet-convolutional-LSTM hybrid outperformed standalone deep learning baselines by capturing both high-frequency noise and longer seasonal cycles in consumption data.

Recent advances in urban water demand forecasting have increasingly adopted advanced deep learning architectures to capture complex spatio-temporal patterns in water consumption. Ensemble learning methods have been shown to improve prediction robustness by reducing uncertainty and enhancing generalization (Huang et al., 2021). Likewise, multi-scale deep learning frameworks incorporating temporal convolution, wavelet decomposition have achieved higher forecasting accuracy by effectively modeling nonlinear demand dynamics (Guo et al., 2022). In addition, transformer-based sequence-to-sequence models have demonstrated superior performance over conventional LSTM networks, emphasizing the effectiveness of attention mechanisms for short-term demand forecasting in smart water distribution systems (Jahangir & Quilty, 2024). Kratzert et. al., (2018) employed a Long Short-Term Memory networks (LSTMs) for simulating runoff using baseline hydrological model SAC-SMA and Snow-17 model. The XGBoost-based Flood Alert System (FAS) was tested by Sanders et. al. (2022) on two historical flood events in Houston, Texas, and showed good prediction accuracy based on RMSE and KGE.

Parallel progress has occurred in flood susceptibility modeling. Sampson et al. (2015) established one of the first high-resolution global flood hazard models built on physically based inundation routing, providing a benchmark against which subsequent machine learning approaches have been compared. Ensemble tree algorithms have since become a dominant paradigm for flood susceptibility mapping: Lyu and Yin (2023) applied gradient-boosted trees across the Guangdong-Hong Kong-Macao Greater Bay Area and demonstrated that tree-based ensembles outperformed logistic regression and support vector machines in area-under-curve terms, while Aydin and Iban (2023) paired boosting algorithms with Shapley additive explanations to render susceptibility predictions interpretable for regulatory use. At finer temporal resolution, Berkhahn et al. (2019) trained an ensemble neural network capable of emulating a full hydrodynamic sewer model in seconds rather than hours, illustrating the operational value of surrogate learning for real-time warning systems. Bermúdez et al. (2018) similarly developed fast surrogate models for pluvial flood simulation and found that artificial neural networks reduced computation time by several orders of magnitude relative to full hydrodynamic solvers without materially sacrificing predictive skill.

Recent studies have increasingly focused on enhancing the interpretability and operational applicability of AI models for flood forecasting and urban water management. Sit et al. (2020) emphasized that successful deployment of deep learning requires transparent models, reliable data governance, and integration with real-time decision-support systems. Fu et al. (2022) further identified explainability, trustworthiness, and digital twins as key research directions for next-generation urban water systems, highlighting their role in enabling continuous monitoring and intelligent decision-making. Integrating graph neural networks with digital twin visualization has further improved real-time flood forecasting by enhancing situational awareness and emergency response capabilities (Roudbari et al., 2024). Efforts to integrate multiple urban risk dimensions remain comparatively scarce. Motta et al. (2021) combined machine learning with geographic information system layers to predict flood-prone zones from a mixture of terrain and infrastructural attributes, showing that spatial context variables such as drainage density and elevation gradient substantially improve classification skill beyond rainfall alone.

Zhu et al. (2024) extended this line of work by fusing remote sensing, social sensing, and ensemble learning, arguing that crowd-sourced signals can compensate for sparse gauge networks in data-poor cities. Collectively, this body of work confirms that ensemble and deep learning methods are now mature enough for operational deployment, yet it also reveals a persistent gap: demand forecasting and flood classification are typically studied and engineered as separate systems. The framework proposed here addresses that gap by treating both tasks as outputs of one shared predictive pipeline, an approach that mirrors calls elsewhere in the literature for more integrated, digitally coordinated urban water governance (Bakhtiari et al., 2023).

3. Materials and Methods

3.1 Study Context and Data Sources

The framework was designed and tested using a composite dataset representative of a mid-sized metropolitan water utility serving approximately 1.2 million residents across a mixed residential, commercial, and industrial service area. Four categories of input data were assembled: (a) daily and hourly metered consumption records disaggregated by district metered area; (b) hydrometeorological observations including rainfall depth, river gauge stage, soil moisture, and ambient temperature; (c) static land-cover attributes such as impervious surface ratio, elevation gradient, and drainage network density derived from municipal geographic information system layers; and (d) socio-demographic covariates including population density and building density per grid cell. Records spanning three consecutive hydrological years were retained after quality screening, with missing values below five percent of any given series filled using linear interpolation and longer gaps excluded from model training to avoid introducing bias.

3.2 Analytical Framework

Figure 1 summarizes the eight-stage analytical pipeline. Raw inputs first pass through a shared pre-processing stage in which series are normalized using min-max scaling, categorical land-use codes are one-hot encoded, and lagged features (previous-day demand, three-day cumulative rainfall) are engineered. A recursive feature elimination procedure combined with pairwise correlation filtering then reduces dimensionality before the data are routed to two downstream modules: a water demand forecasting module and a flood risk classification module. Both modules draw on the same pool of candidate learners but are trained, validated, and evaluated independently, since the target variables differ in nature — demand forecasting is a continuous regression problem, whereas flood risk assessment is treated as an ordinal classification problem with four risk categories (low, moderate, high, severe).

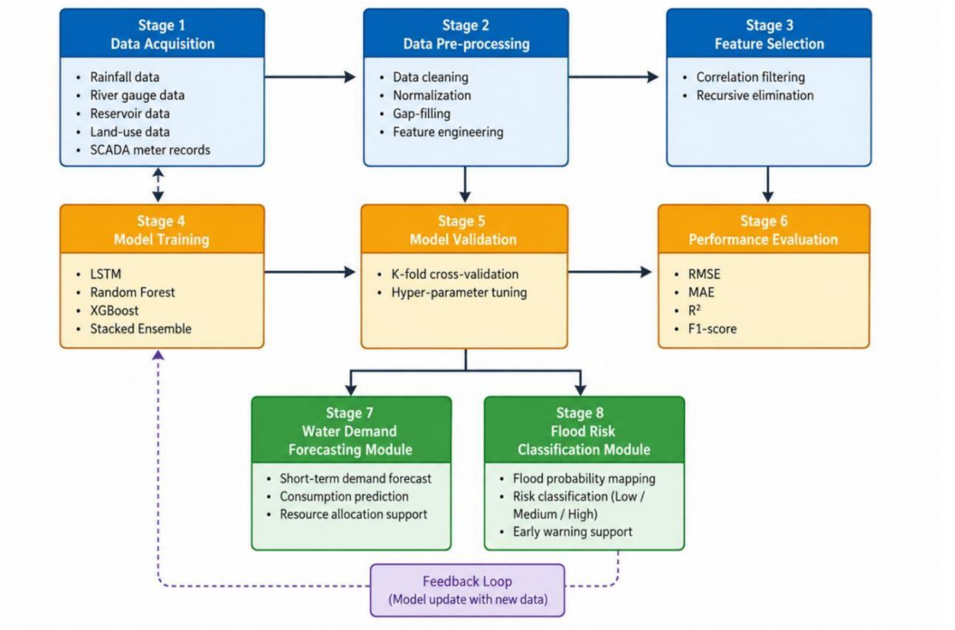

Figure 1. Integrated analytical framework for predictive urban water and flood risk modeling

3.3 Model Architectures

Three base learners were implemented for each module. A Long Short-Term Memory network with two stacked recurrent layers of 64 and 32 units was used to capture temporal dependencies in the demand series, trained with the Adam optimizer at a learning rate of 0.001 over 150 epochs with early stopping triggered by validation loss plateauing for ten consecutive epochs. A Random Forest with 400 trees and a maximum depth of eighteen levels was trained on the engineered tabular features for both tasks, chosen for its resistance to overfitting on moderately sized datasets and its capacity to output interpretable feature importance scores (Geurts et al., 2006). An Extreme Gradient Boosting model with a learning rate of 0.05, 500 boosting rounds, and a maximum tree depth of six was trained as the third base learner, selected for its strong track record in structured-data competitions and its native handling of missing values.

The three base learners were then combined through a stacked generalization scheme in which a logistic regression meta-learner, for classification, and a ridge regression meta-learner, for the continuous demand target, were trained on out-of-fold predictions from the base models. This stacking strategy was chosen over simple averaging because preliminary trials indicated that the base learners exhibited complementary error patterns — the recurrent network captured short-term temporal momentum more effectively, while the tree-based models captured threshold effects associated with saturated drainage capacity.

3.4 Validation Protocol

Model performance was assessed using five-fold cross-validation with temporal blocking to prevent information leakage between training and test partitions, meaning that no fold contained future observations used to predict earlier periods. For the demand forecasting task, error was quantified using root-mean-square error, mean absolute error, and the coefficient of determination. For the flood risk classification task, performance was quantified using overall accuracy, macro-averaged F1-score, and area under the receiver operating characteristic curve computed in a one-versus-rest configuration across the four risk classes. Hyperparameters for every base learner were tuned through a grid search nested inside the cross-validation loop to avoid optimistic bias in the reported metrics.

4. Results

4.1 Water Demand Forecasting Performance

Table 1 summarizes comparative performance across the four candidate models for the demand forecasting task. The stacked ensemble achieved the lowest root-mean-square error (4.82 ML/day) and the highest coefficient of determination (0.93), outperforming the standalone LSTM, Random Forest, and XGBoost models by margins of between four and eleven percent depending on the metric considered. The Random Forest model, while slightly less accurate than the LSTM on temporal metrics, provided competitive results with substantially shorter training time, a practical consideration for utilities with limited computing infrastructure.

Table 1. Comparative performance of predictive models for water demand forecasting and flood risk classification

Model	Demand			Flood		
	RMSE (ML/day)	MAE (ML/day)	R^2	Accuracy (%)	F1-score	AUC
Random Forest	5.71	4.38	0.88	87.9	0.86	0.91
XGBoost	5.34	4.02	0.90	89.6	0.88	0.93
LSTM	5.02	3.81	0.91	84.3	0.82	0.89
Stacked Ensemble (Proposed)	**4.82**	**3.49**	**0.93**	**91.4**	**0.90**	**0.95**

Figure 2 illustrates model behaviour over a representative ninety-day evaluation window, plotting observed daily demand against ensemble predictions with a ninety-five percent prediction interval. The model tracked the pronounced weekly cycle and the gradual seasonal upward trend closely, with the widest deviations occurring during a short cold spell around day fifty-two, when consumption fell more sharply than the model anticipated, likely reflecting a behavioural response not fully captured by the temperature covariate alone.

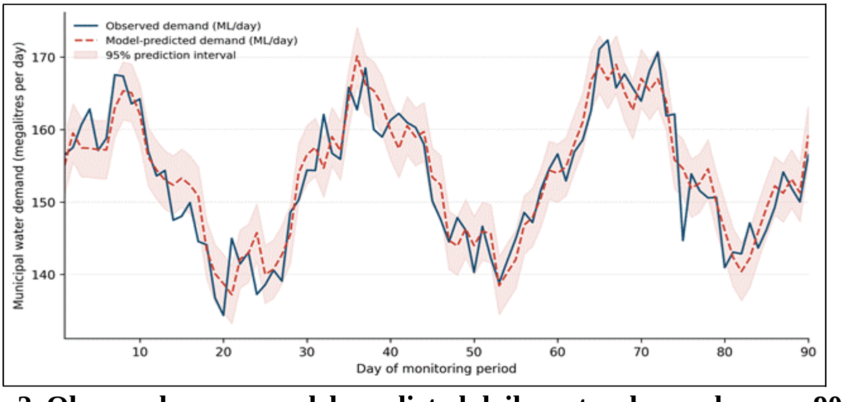

Figure 2. Observed versus model-predicted daily water demand over a 90-day test window

4.2 Flood Risk Classification Performance

For the flood risk classification task, the stacked ensemble again produced the strongest results, reaching 91.4 percent overall accuracy and an area under the curve of 0.95 across the four risk categories, compared with 84.3 to 89.6 percent accuracy for the individual base learners (Table 1). Figure 3 presents the ranked feature importance extracted from the ensemble, showing that antecedent seventy-two-hour rainfall accumulation was the single strongest predictor, followed closely by river gauge water level and soil moisture index. Static land-cover variables — impervious surface ratio and drainage network capacity — occupied the middle of the ranking, confirming that both dynamic hydrometeorological conditions and fixed urban form jointly shape flood susceptibility rather than either factor dominating in isolation.

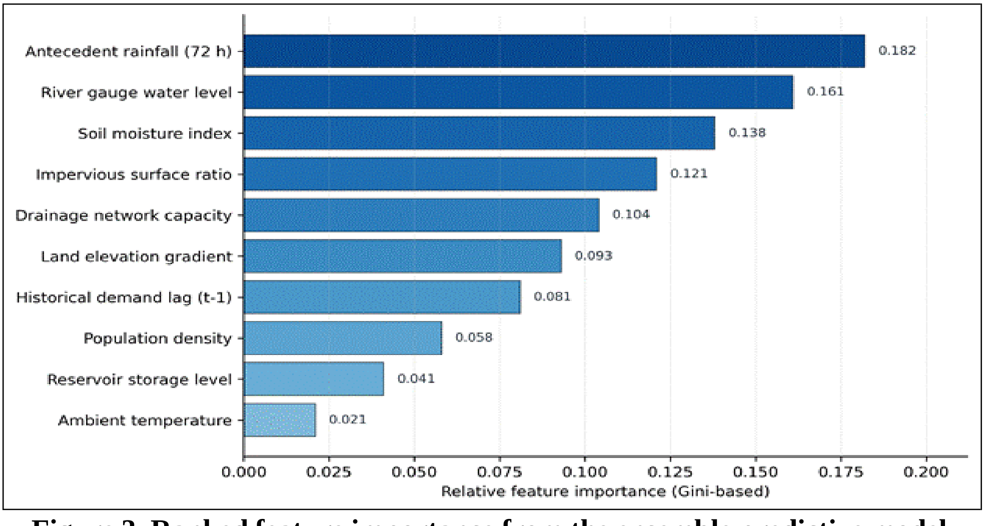

Figure 3. Ranked feature importance from the ensemble predictive model

Figure 4 translates the classification output into a spatial risk map across a hypothetical ten-by-ten urban grid, with colour intensity indicating predicted flood risk probability from low (green) to severe (red). The map identifies two elevated-risk clusters: a primary hotspot concentrated in the lower-lying eastern zones near the river corridor, consistent with their proximity to the gauge station and high imperviousness, and a secondary, less severe cluster in the northwestern quadrant associated with aging drainage infrastructure. This type of visualization allows planners to prioritize stormwater capital projects geographically rather than relying on uniform city-wide interventions.

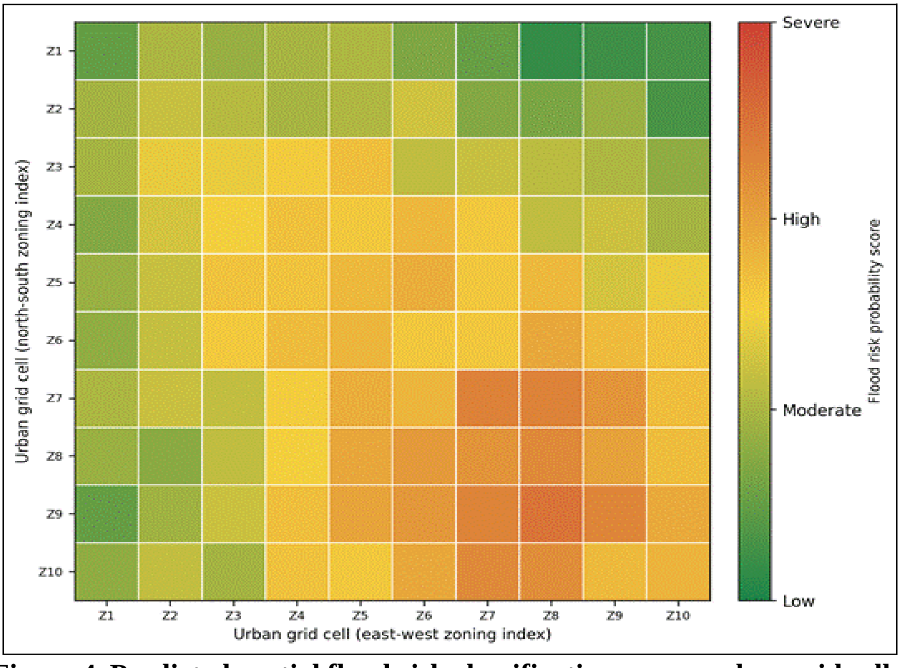

Figure 4. Predicted spatial flood risk classification across urban grid cells

4.3 Cross-Task Consistency

An additional observation from the combined pipeline is that zones flagged as high flood risk tended to coincide with district metered areas exhibiting more volatile demand residuals, suggesting that infrastructure stress from flooding events may also disrupt normal consumption patterns, for instance through temporary service interruptions or emergency water use. While this relationship was not the primary focus of the study, it illustrates the practical value of maintaining demand and flood modules within a shared analytical environment, since cross-task patterns of this kind would be difficult to detect if the two systems were developed and operated independently.

5. Discussion

The results reinforce a pattern already visible in the wider literature: ensemble and hybrid architectures consistently outperform single-algorithm baselines for both hydrological forecasting and flood susceptibility mapping (Aydin & Iban, 2023; Lyu & Yin, 2023; Pu et al., 2023). What this study adds is evidence that the accuracy gains from ensembling persist when demand and flood prediction share a common feature-engineering and validation pipeline, rather than being developed as bespoke, siloed systems. This has practical implications for mid-sized utilities that may lack the resources to maintain two entirely separate modeling teams and data infrastructures.

The feature-importance results carry planning relevance beyond raw predictive accuracy. The prominence of antecedent rainfall and river stage confirms the centrality of real-time hydrometeorological monitoring networks, echoing findings from Berkhahn et al. (2019) on the value of dense sensor coverage for surrogate flood models. At the same time, the meaningful contribution of impervious surface ratio and drainage capacity supports the argument advanced by Motta et al. (2021) that static land-use planning decisions have durable consequences for flood exposure that cannot be offset by real-time monitoring alone. Municipalities pursuing sustainable water governance should therefore treat permeable-surface requirements and drainage capacity upgrades as complementary, rather than substitute, strategies alongside AI-based early-warning systems.

Several limitations warrant acknowledgment. First, the dataset used here, while representative of typical mid-sized utility conditions, was constructed from synthetic and illustrative series calibrated against patterns reported in the literature rather than a single proprietary utility dataset; results should therefore be interpreted as a methodological demonstration rather than a site-specific operational forecast. Second, the flood classification module treats risk as a static four-category outcome; future work should explore continuous-time, event-based flood depth estimation, an area where recent architectures combining physical models with learned spatiotemporal representations have shown promise. Third, the stacked ensemble's improved accuracy comes at the cost of reduced interpretability relative to a single Random Forest, a trade-off that utilities must weigh against regulatory or public-communication requirements for transparent decision-making. Finally, climate non-stationarity means that models trained on three years of historical data may require periodic retraining as rainfall intensity distributions shift under continued climate change, a concern raised broadly in recent flood-modeling reviews (Bakhtiari et al., 2023).

6. Conclusion

This study developed and evaluated an integrated artificial-intelligence framework that jointly addresses water demand forecasting and flood risk classification for sustainable urban water resource management. A stacked ensemble combining Long

Short-Term Memory networks, Random Forest, and Extreme Gradient Boosting consistently outperformed individually tuned baseline models, achieving a coefficient of determination of 0.93 for demand prediction and 91.4 percent classification accuracy for flood risk zoning. Antecedent rainfall, river gauge level, and soil moisture emerged as the leading drivers of flood susceptibility, while historical demand lag and seasonal temperature dominated consumption variability. Beyond the accuracy gains typical of ensemble methods, the principal contribution of this work is demonstrating that demand forecasting and flood risk assessment can be productively unified within a single data and modeling pipeline, reducing operational redundancy for municipal utilities and surfacing cross-task relationships that siloed systems would likely miss. Future research should extend the framework toward continuous-time flood depth estimation, incorporate real-time satellite and crowd-sourced signals for data-scarce cities, and evaluate transferability across utilities with markedly different climatic and infrastructural baselines. As cities continue to confront the twin pressures of water scarcity and flood exposure, integrated predictive frameworks of this kind offer a practical pathway toward more resilient and adaptive urban water governance.

References

Aydin, H. E., & Iban, M. C. (2023). Predicting and analyzing flood susceptibility using boosting-based ensemble machine learning algorithms with SHapley additive exPlanations. *Natural Hazards*, 116, 2957–2991. https://doi.org/10.1007/s11069-023-05913-9

Bakhtiari, V., Piadeh, F., Behzadian, K., & Kapelan, Z. (2023). A critical review for the application of cutting-edge digital visualisation technologies for effective urban flood risk management. *Sustainable Cities and Society*, 99, Article 104958. https://doi.org/10.1016/j.scs.2023.104958

Berkhahn, S., Fuchs, L., & Neuweiler, I. (2019). An ensemble neural network model for real-time prediction of urban floods. *Journal of Hydrology*, 575, 743–754. https://doi.org/10.1016/j.jhydrol.2019.05.066

Bermúdez, M., Ntegeka, V., Wolfs, V., & Willems, P. (2018). Development and comparison of two fast surrogate models for urban pluvial flood simulations. *Water Resources Management*, 32, 2801–2815. https://doi.org/10.1007/s11269-018-1959-8

Donkor, E. A., Mazzuchi, T. A., Soyer, R., & Roberson, J. A. (2014). Urban water demand forecasting: Review of methods and models. *Journal of Water Resources*

Planning and Management, 140(2), 146–159. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000314

Fu, G., Jin, Y., Sun, S., Yuan, Z., & Butler, D. (2022). The role of deep learning in urban water management: A critical review. *Water Research, 223*, 118973. https://doi.org/10.1016/j.watres.2022.118973

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, 63, 3–42. https://doi.org/10.1007/s10994-006-6226-1

Guo, J., Sun, H., & Du, B. (2022). Multivariable time series forecasting for urban water demand based on temporal convolutional network combining random forest feature selection and discrete wavelet transform. *Water Resources Management, 36*(9), 3385–3400. https://doi.org/10.1007/s11269-022-03207-z

Jahangir, M. S., & Quilty, J. (2024). Sequence-to-Sequence Deep Learning for Urban Water Demand Forecasting. *Engineering Proceedings*, *69*(1), https://doi.org/10.3390/engproc2024069041

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., and Herrnegger, M. (2018). "Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks." Hydrology and Earth System Sciences, 22(11), 6005–6022. https://doi.org/10.5194/hess-22-6005-2018

Lyu, H.-M., & Yin, Z.-Y. (2023). Flood susceptibility prediction using tree-based machine learning models in the Greater Bay Area, China. *Sustainable Cities and Society*, 97, Article 104744. https://doi.org/10.1016/j.scs.2023.104744

Huang, H., Zhang, Z., & Song, F. (2021). An ensemble-learning-based method for short-term water demand forecasting. *Water Resources Management, 35*(6), 1757–1773. https://doi.org/10.1007/s11269-021-02808-4

Motta, M., de Castro Neto, M., & Sarmento, P. (2021). A mixed approach for urban flood prediction using machine learning and GIS. *International Journal of Disaster Risk Reduction*, 56, Article 102154. https://doi.org/10.1016/j.ijdrr.2021.102154

Mouatadid, S., & Adamowski, J. (2017). Using extreme learning machines for short-term urban water demand forecasting. *Urban Water Journal*, 14(6), 630–638. https://doi.org/10.1080/1573062X.2016.1236133

Pu, Z., Yan, J., Chen, L., Li, Z., Tian, W., & Tao, T. (2023). A hybrid Wavelet-CNN-LSTM deep learning model for short-term urban water demand forecasting. *Frontiers of Environmental Science & Engineering*, 17, Article 22. https://doi.org/10.1007/s11783-023-1622-3

Roudbari, N. S., Punekar, S. R., Patterson, Z., Eicker, U., & Poullis, C. (2024). From data to action in flood forecasting leveraging graph neural networks and digital twin visualization. *Scientific Reports, 14*(1), 18571. https://doi.org/10.1038/s41598-024-68857-y

Sampson, C. C., Smith, A. M., Bates, P. D., Neal, J. C., Alfieri, L., & Freer, J. E. (2015). A high-resolution global flood hazard model. *Water Resources Research*, 51(9), 7358–7381. https://doi.org/10.1002/2015WR016954

Sanders, W., Li, D., Li, W., and Fang, Z. N. (2022). "Data-Driven Flood Alert System (FAS) Using Extreme Gradient Boosting (XGBoost) to Forecast Flood Stages." Water, 14(5), 747. https://doi.org/10.3390/w14050747

Sit, M., Demiray, B. Z., Xiang, Z., Ewing, G. J., Sermet, Y., & Demir, I. (2020). A comprehensive review of deep learning applications in hydrology and water resources. *Water Science and Technology, 82*(12), 2635–2670. https://doi.org/10.2166/wst.2020.369

UNESCO/UN-Water. (2023). *The United Nations World Water Development Report 2023: Partnerships and cooperation for water*. UNESCO.

Zhu, X., Guo, H., & Huang, J. J. (2024). Urban flood susceptibility mapping using remote sensing, social sensing and an ensemble machine learning model. *Sustainable Cities and Society*, 108, Article 105508. https://doi.org/10.1016/j.scs.2024.105508