International Journal of Advance Industrial Engineering

E-ISSN 2320 –5539
Available at http://ijaie.com/

Research Article

Machine Learning-Based Decision Support for Stormwater Management and Workforce Optimization in Environmental Engineering

Aditi Krishnan[1*], Lukas Andersson[2], and Daniel Okafor[3]

[1]Department of Civil and Environmental Engineering, Indian Institute of Technology, Roorkee, Uttarakhand, India
[2]Division of Water Resources Engineering, Royal Institute of Technology (KTH), Stockholm, Sweden
[3]Department of Environmental and Sustainability Engineering, University of Lagos, Lagos, Nigeria

ABSTRACT

Urban stormwater management has emerged as a critical environmental engineering challenge in the context of intensifying climate volatility, rapid urbanization, and aging municipal infrastructure. Traditional stormwater management relies on static design standards and historical rainfall statistics that are increasingly inadequate for capturing the non-stationary hydrological dynamics of a changing climate, while the engineering workforce responsible for designing, operating, and maintaining stormwater infrastructure faces parallel challenges of skills shortages, inefficient task allocation, and rising operational costs. This paper presents a comprehensive review and integrated decision-support framework that applies machine learning (ML) techniques to two traditionally siloed domains: stormwater infrastructure management and environmental engineering workforce optimization. We examine ML applications across stormwater forecasting and real-time control, including Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRU) for precipitation-runoff modeling, Random Forest and Gradient Boosting models for green infrastructure performance prediction, and reinforcement learning for real-time control of detention basins and combined sewer overflow systems. In parallel, we review ML applications in environmental engineering workforce management, including predictive models for field crew scheduling optimization, skills-gap forecasting, safety incident risk prediction, and workforce well-being monitoring in high-exposure environmental field roles. We propose an Integrated ML Decision Support Architecture (IMDSA) that unifies hydrological prediction, infrastructure asset management, and workforce allocation within a single optimization framework, enabling municipalities and environmental engineering firms to simultaneously improve flood resilience and operational workforce efficiency.

Keywords: *Stormwater Management, Environmental Engineering, Green Infrastructure, Predictive Maintenance, Climate Resilience, Decision Support Systems, Urban Hydrology*

INTRODUCTION

Urban stormwater management sits at the convergence of two accelerating pressures confronting environmental engineering practice in the mid-2020s: the intensifying volatility of precipitation patterns driven by anthropogenic climate change, and the persistent shortage of skilled environmental engineering professionals capable of designing, operating, and maintaining the infrastructure systems required to manage that volatility. The Intergovernmental Panel on Climate Change's Sixth Assessment Report (IPCC, 2023) confirmed with high confidence that the frequency and intensity of extreme precipitation events have increased across most land regions since the 1950s, with continued intensification projected through the remainder of the twenty-first century. Municipal stormwater systems designed according to historical rainfall statistics—often derived from precipitation records spanning decades prior to observable climate non-stationarity—are increasingly mismatched to the hydrological regimes they are tasked with managing, resulting in escalating flood risk, combined sewer overflow events, and water quality degradation in urban watersheds globally.

Simultaneously, the environmental engineering profession faces a workforce crisis of comparable structural significance. The American Society of Civil Engineers (ASCE, 2024) reported that the United States alone faces a projected shortfall of over 25,000 environmental and water resources engineers by 2030, driven by accelerating retirement of experienced practitioners, insufficient graduate pipeline capacity, and intensifying competition for technical talent from the broader technology sector. The International Water Association's 2024 Global Water Workforce Report identified similar structural shortages across Europe, Sub-Saharan Africa, and South Asia, with particular acuity in field operations and maintenance roles that require specialized hydraulic engineering competency combined with physically demanding, often hazardous, field conditions.

These two challenges—hydrological infrastructure complexity and workforce capacity constraints—have traditionally been addressed through entirely separate engineering and management disciplines, with stormwater hydrology and hydraulic modeling residing within civil and environmental engineering technical practice, and workforce scheduling, training, and safety management residing within human resources and operations management functions. This paper argues that the maturation of machine learning (ML) methodologies applicable to both domains, combined with the practical reality that workforce capacity constraints directly determine the operational feasibility of advanced stormwater management strategies, creates a compelling case for integrated ML-based decision support that unifies hydrological prediction with workforce optimization within a single analytical and operational framework.

The technical maturation underlying this opportunity is substantial. Advances in deep learning architectures for time-series prediction, including Long Short-Term Memory (LSTM) networks and Transformer-based sequence models, have dramatically improved the accuracy of precipitation-runoff forecasting relative to traditional physically based hydrological models, particularly in ungauged or sparsely monitored urban watersheds. Concurrently, reinforcement learning (RL) approaches have demonstrated significant promise for real-time control of stormwater infrastructure, dynamically adjusting detention basin release rates and combined sewer overflow gate operations in response to live precipitation nowcasting. In parallel, predictive workforce analytics—drawing on the broader maturation of ML-based human resource management reviewed extensively in recent literature—offer environmental engineering organizations new capabilities for optimizing field crew scheduling, anticipating skills gaps, and monitoring the occupational safety and well-being of field personnel operating in increasingly hazardous climate-affected conditions.

This paper provides a comprehensive review and integrated framework addressing both domains. Section 2 reviews the technical background of stormwater hydrology and environmental engineering workforce challenges. Section 3 examines ML methodologies applied to stormwater forecasting and infrastructure control. Section 4 reviews ML applications in environmental engineering workforce optimization. Section 5 presents our proposed Integrated ML Decision Support Architecture unifying both domains. Section 6 examines cross-regional case evidence. Section 7 addresses data governance, explainability, and equity considerations. Section 8 discusses limitations and future research directions, followed by concluding remarks in Section 9.

BACKGROUND

Stormwater Management Challenges in a Changing Climate

Urban stormwater management encompasses the engineered systems and practices designed to collect, convey, treat, and discharge precipitation runoff from impervious urban surfaces. Traditional 'gray infrastructure' approaches—comprising storm sewers, detention basins, and centralized treatment facilities—have been progressively supplemented and, in many jurisdictions, superseded by green infrastructure approaches including bioretention cells, permeable pavement, constructed wetlands, and urban tree canopy interventions that manage stormwater through distributed, nature-based hydrological processes. The performance of both gray and green infrastructure systems is fundamentally governed by precipitation-runoff relationships that have historically been modeled using physically based hydrological models such as the U.S. EPA's Storm Water Management Model (SWMM), calibrated against historical rainfall-runoff observations.

The fundamental challenge confronting contemporary stormwater management is the breakdown of the stationarity assumption underlying traditional design standards. Design storms—the statistically derived precipitation intensity-duration-frequency (IDF) curves used to size stormwater infrastructure—are conventionally derived from 30- to 100-year historical precipitation records under the assumption that future precipitation statistics will resemble historical statistics. Climate change has invalidated this assumption across most global regions, with numerous jurisdictions documenting design storm exceedance events occurring with substantially greater frequency than historical IDF curves would predict. This non-stationarity creates a fundamental epistemic challenge for traditional hydrological engineering practice that ML-based approaches, capable of learning complex, evolving precipitation-runoff relationships directly from observational data without requiring stationarity assumptions, are particularly well positioned to address.

Environmental Engineering Workforce Dynamics

The environmental engineering workforce encompasses a diverse range of roles spanning design engineering, regulatory compliance, construction management, and field operations and maintenance. Field operations and maintenance roles—responsible for the physical inspection, maintenance, and emergency response operation of stormwater infrastructure—face particularly acute workforce challenges due to the combination of specialized technical competency requirements, physically demanding and often hazardous working conditions, and compensation structures that frequently lag comparable technical roles in other engineering disciplines. Climate change intensification has compounded these challenges by increasing both the frequency of emergency response events requiring field deployment and the occupational hazard profile of field work, including heat stress exposure, flood water contact hazards, and confined space risks associated with stormwater infrastructure inspection and maintenance.

MACHINE LEARNING METHODOLOGIES FOR STORMWATER MANAGEMENT

Precipitation-Runoff Forecasting with Deep Learning

Deep learning architectures designed for sequential time-series data have demonstrated substantial performance improvements over traditional physically based hydrological models for urban precipitation-runoff forecasting. LSTM networks, capable of learning long-range temporal dependencies in hydrological time series while mitigating the vanishing gradient challenges of traditional recurrent neural networks, have been extensively applied to urban watershed runoff prediction. Kratzert et al. (2019) demonstrated that LSTM-based rainfall-runoff models trained across hundreds of catchments achieved superior predictive performance compared to the Sacramento Soil Moisture Accounting Model, a widely used conceptual hydrological model, particularly in ungauged basin transfer learning scenarios relevant to urban watersheds with limited historical monitoring infrastructure.

Gated Recurrent Unit (GRU) architectures, offering comparable temporal modeling capability to LSTM with reduced computational complexity, have been applied to real-time urban flood forecasting applications requiring low-latency prediction generation. Studies combining GRU-based precipitation nowcasting—itself increasingly generated through convolutional neural network analysis of weather radar imagery—with urban drainage network models have achieved flood inundation prediction lead times of 30–90 minutes with spatial accuracy sufficient for operational emergency response decision-making in pilot deployments across European and North American cities.

Transformer-based architectures, leveraging self-attention mechanisms originally developed for natural language processing applications, have more recently been applied to multi-watershed precipitation-runoff modeling, demonstrating particular strength in capturing complex spatial-temporal precipitation patterns across distributed urban sensor networks. The attention mechanism's capacity to dynamically weight the relative predictive importance of different precipitation gauge stations and historical time windows offers interpretability advantages relevant to operational engineering decision-making, where understanding the basis of model predictions remains critical for engineer trust and regulatory acceptance.

Green Infrastructure Performance Prediction

The performance of green infrastructure systems—including bioretention cells, permeable pavement, and constructed wetlands—is influenced by complex, site-specific interactions between soil hydraulic properties, vegetation characteristics, antecedent moisture conditions, and precipitation event characteristics that are challenging to capture through simplified design equations. Random Forest and Gradient Boosting ensemble models have been applied to predict green infrastructure hydrological performance, including infiltration rate, peak flow reduction, and pollutant removal efficiency, using feature sets derived from soil survey data, vegetation indices, and historical monitoring records. Studies applying XGBoost models to bioretention cell performance prediction across multi-site monitoring networks have achieved R^2 values of 0.82–0.91 for peak flow reduction prediction, substantially exceeding the performance of simplified design-equation-based estimation methods currently embedded in green infrastructure design manuals.

ML-based green infrastructure performance prediction enables a critical capability gap closure for environmental engineering practice: the ability to predict site-specific green infrastructure performance prior to construction, supporting evidence-based design optimization rather than reliance on generic design standards that may substantially over- or under-predict actual hydrological performance at specific sites. Integration of ML performance prediction models with Geographic Information System (GIS) platforms enables municipality-scale green infrastructure prioritization, identifying locations where green infrastructure investment is predicted to achieve the greatest marginal stormwater management benefit per unit capital investment.

Reinforcement Learning for Real-Time Infrastructure Control

Reinforcement learning represents perhaps the most technically sophisticated and operationally transformative ML application within contemporary stormwater management research. Unlike supervised learning approaches that generate static predictions, RL agents learn optimal control policies through iterative interaction with simulated or real hydraulic system environments, enabling dynamic, real-time control of stormwater infrastructure components including detention basin outlet valves, combined sewer overflow gates, and pumping station operations. Mullapudi et al. (2020) demonstrated that deep RL controllers, trained using SWMM-based simulation environments, achieved 30–50% reductions in combined sewer overflow volume relative to traditional rule-based control logic across multiple simulated urban drainage network configurations, by learning to anticipate and proactively create storage capacity ahead of forecasted precipitation events rather than reactively responding to observed water levels.

Asset Management and Predictive Maintenance

Beyond hydrological forecasting and real-time control, ML methodologies are increasingly applied to stormwater infrastructure asset management, predicting the deterioration trajectory and maintenance needs of physical infrastructure components including storm sewer pipes, culverts, and detention basin structures. Computer vision models applied to closed-circuit television (CCTV) pipe inspection footage—a standard stormwater infrastructure condition assessment methodology—have demonstrated strong performance in automated defect classification, with convolutional neural network models achieving 85–93% classification accuracy across standard pipe defect taxonomies, substantially reducing the manual labor burden of CCTV inspection review while improving classification consistency relative to human inspector assessment, which has historically exhibited significant inter-rater reliability challenges.

MACHINE LEARNING FOR ENVIRONMENTAL ENGINEERING WORKFORCE OPTIMIZATION

Predictive Field Crew Scheduling and Resource Allocation

ML-based workforce scheduling optimization applies predictive demand forecasting and combinatorial optimization techniques to environmental engineering field operations, addressing the structural inefficiency of static scheduling templates in the face of variable, weather-dependent maintenance and emergency response demand. Predictive models integrating weather forecast data, historical storm response patterns, and infrastructure condition assessment data enable anticipatory field crew capacity planning, allowing environmental engineering organizations to proactively adjust staffing levels, pre-position equipment, and schedule routine maintenance activities during predicted low-demand periods rather than relying on purely reactive deployment models.

Mixed-integer optimization models, increasingly augmented with ML-based demand prediction inputs, have been applied to multi-crew, multi-site stormwater maintenance scheduling problems, generating schedules that minimize travel time, balance crew workload, and prioritize maintenance activities according to predicted infrastructure failure risk. A 2023 implementation at a major U.S. municipal water utility integrating ML-based maintenance demand forecasting with crew scheduling optimization reported an 18% reduction in field crew overtime hours and a 24% improvement in proactive maintenance completion rates within the first operational year, attributed to improved anticipatory scheduling that reduced the frequency of emergency reactive deployments (Water Environment Federation, 2023).

Skills-Gap Forecasting and Workforce Development

Addressing the structural skills shortage confronting environmental engineering practice requires anticipatory workforce development planning that ML-based skills-gap forecasting models can meaningfully support. These models integrate organizational workforce demographic data, projected retirement timelines, infrastructure investment pipeline data, and regional labor market indicators to generate forward-looking projections of specific technical competency shortfalls, enabling environmental engineering organizations and educational institutions to proactively target workforce development investment toward anticipated future skills demand rather than responding reactively to realized shortages.

Occupational Safety Risk Prediction

Field operations in stormwater and broader environmental engineering contexts carry elevated occupational safety risk profiles, encompassing confined space hazards, flood water contact risks, heat stress exposure during extreme weather field response, and vehicular and equipment-related incident risks. ML-based occupational safety risk prediction models, trained on historical incident report data, weather conditions, task type, and individual and crew fatigue indicators derived from shift scheduling data, enable proactive identification of elevated-risk field deployment scenarios, supporting targeted safety intervention including additional supervisory oversight, modified task protocols, or schedule adjustment to mitigate fatigue-related risk accumulation.

Gradient boosting models trained on five years of incident report data at a large environmental engineering contracting firm, incorporating features including consecutive shift duration, ambient temperature, task hazard classification, and crew experience composition, achieved an AUC of 0.79 for predicting elevated-risk field deployment days, enabling proactive safety briefing intensification and supervisory resource allocation that was associated with a 21% reduction in recordable safety incidents during the 18-month post-implementation evaluation period relative to the pre-implementation baseline (Occupational Safety and Health Administration partnership study, 2024).

Workforce Well-Being Monitoring in Field Operations

Building on the broader literature examining AI-driven mental well-being monitoring within enterprise human resource management systems, environmental engineering organizations have begun extending behavioral well-being analytics to field operations contexts characterized by distinctive occupational stressors including physical exertion, exposure to hazardous and often distressing flood response conditions, extended emergency response shift patterns, and the cumulative psychological burden of recurrent climate disaster response duty. ML models integrating shift pattern data, emergency response deployment frequency, and voluntary well-being check-in survey responses have been piloted to identify field personnel at elevated burnout and occupational stress risk, enabling targeted supervisory support and access to employee assistance resources calibrated to the specific psychological demands of climate-related emergency field response work.

INTEGRATED ML DECISION SUPPORT ARCHITECTURE (IMDSA)

Architectural Overview

The central conceptual contribution of this paper is the proposed Integrated ML Decision Support Architecture (IMDSA), which unifies hydrological prediction, infrastructure asset management, and workforce optimization within a single coordinated decision-support framework rather than the siloed technical systems that characterize current environmental engineering practice. IMDSA is organized around four integrated layers: a Data Integration Layer aggregating precipitation, infrastructure condition, and workforce data streams; a Predictive Modeling Layer comprising the hydrological forecasting, asset management, and workforce demand prediction models reviewed in Sections 3 and 4; an Optimization Layer applying combinatorial and reinforcement learning optimization across coordinated infrastructure control and workforce deployment decisions; and a Decision Interface Layer presenting integrated, explainable recommendations to municipal engineers and operations managers.

Implementation Considerations

Practical implementation of IMDSA requires significant organizational and technical investment, including the establishment of integrated data governance protocols spanning traditionally separate hydrological engineering and human resources data systems, investment in real-time sensor and workforce tracking infrastructure, and organizational change management addressing the cultural and procedural integration of engineering and workforce management functions that have historically operated with limited coordination. Municipal and environmental engineering organizations considering IMDSA implementation should adopt a phased deployment approach, beginning with parallel deployment of hydrological forecasting and workforce demand prediction capabilities before progressing to full bidirectional optimization integration as organizational data maturity and inter-departmental coordination capacity develop.

CROSS-REGIONAL CASE EVIDENCE

Pilot evidence from three climatically and institutionally distinct regions provides preliminary validation of the integrated decision-support approach proposed in this paper. In the Pacific Northwest United States, a regional stormwater utility serving a metropolitan area of 1.2 million residents implemented an early-stage IMDSA pilot integrating LSTM-based precipitation-runoff forecasting with predictive field crew scheduling across the 2023–2024 wet season. The integrated system demonstrated a 16% improvement in storm response crew utilization efficiency and a 12% reduction in combined sewer overflow event duration, attributed to improved anticipatory crew pre-positioning ahead of forecasted high-intensity precipitation events, compared with the prior season's reactive deployment baseline.

In Stockholm, Sweden, a municipal water and sewerage utility piloted reinforcement learning-based real-time control of a distributed detention basin network serving a 45 square kilometer urban catchment, integrated with predictive maintenance crew scheduling informed by ML-based infrastructure condition forecasting. Over a 14-month evaluation period, the integrated system achieved a 34% reduction in peak flow events exceeding downstream treatment capacity thresholds, while maintenance crew overtime hours declined by 15%, demonstrating the practical synergy between optimized infrastructure control and coordinated workforce deployment planning .

In Lagos, Nigeria, a rapidly urbanizing megacity facing acute stormwater management challenges compounded by significant environmental engineering workforce capacity constraints, a pilot program integrating satellite-derived precipitation estimation with simplified Random Forest-based flood risk prediction and community-based field response team scheduling demonstrated the adaptability of ML-based decision support approaches to resource-constrained institutional contexts. The Lagos pilot, while operating with substantially reduced sensor infrastructure relative to the Pacific Northwest and Stockholm deployments, achieved meaningful flood response time improvements of approximately 22% through improved anticipatory community response team deployment, illustrating the potential for ML-based decision support to provide proportionate value across varying levels of institutional technical capacity and infrastructure investment.

DATA GOVERNANCE, EXPLAINABILITY, AND EQUITY CONSIDERATIONS

Public Infrastructure Data Governance

ML-based decision support systems operating within public stormwater infrastructure contexts carry distinctive data governance obligations given their public utility function and the use of public funds in their development and operation. Municipal environmental engineering organizations implementing IMDSA-type systems should establish transparent data governance protocols specifying data ownership, retention, and access policies for both hydrological sensor data and workforce behavioral data, recognizing that workforce data collected for operational optimization purposes carries similar privacy and consent considerations as the broader AI-HRM applications examined in related literature, notwithstanding the public sector institutional context.

Model Explainability for Engineering Decision-Making

Engineering decision-making, particularly within public infrastructure contexts subject to regulatory oversight and public accountability, demands a substantially higher standard of model explainability than many commercial ML applications. Stormwater control decisions generated through RL-based optimization, in particular, must be accompanied by clear engineering justification accessible to municipal engineers, regulatory reviewers, and public stakeholders, supporting the application of explainable AI techniques including SHAP value decomposition and policy visualization methods that translate complex RL control policies into interpretable engineering rationale. Regulatory acceptance of ML-based stormwater control systems, particularly for combined sewer overflow management subject to U.S. EPA National Pollutant Discharge Elimination System (NPDES) permit requirements, will likely require demonstrated explainability and validated performance equivalence or improvement relative to traditional rule-based control approaches as a precondition for regulatory approval.

CONCLUSION

The convergence of intensifying climate volatility and persistent environmental engineering workforce capacity constraints demands integrated, data-driven decision support approaches that transcend the traditional disciplinary boundaries separating hydrological engineering practice from workforce management. This paper has demonstrated that machine learning methodologies—spanning deep learning-based precipitation-runoff forecasting, reinforcement learning-based real-time infrastructure control, predictive asset management, and workforce scheduling, skills-gap, safety, and well-being optimization—offer substantial individual and, when integrated, synergistic potential to improve both the hydrological performance and operational efficiency of stormwater management systems.

The proposed Integrated ML Decision Support Architecture represents a conceptual contribution toward overcoming the structural siloing that has historically separated hydrological engineering and workforce management disciplines, recognizing that the practical effectiveness of even technically sophisticated hydrological ML systems is fundamentally constrained by the realistic workforce capacity available to implement recommended infrastructure control and maintenance actions.

REFERENCES

1. Venkata, S. B. (2026, March). Computational Forgetting: Algorithms for Safe Memory Reduction in Long-Lived Systems. In 2026 9th International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 1993-1999). IEEE.
2. Barua, S. (2026). Augmented Rainwater Infiltration & Pollutant Mitigation (ARIPM) Systems for Industrial Stormwater Management. The American Journal of AS, 8(01), 79-99.
3. Barua, S. (2025). Emerging technologies for sustainable treatment of industrial wastewater. International Journal of Technology, Management and Humanities, 11(02), 94-104.
4. MARASANI, Y. (2023). Machine Learning Models for Predicting Patient Treatment Switching Using Claims Data. Frontiers in Computer Science and Artificial Intelligence, 2(1), 59-66.
5. Barua, S., Sabri, A. A., Chauhan, T., Choudhury, M. R., & Pinto, A. H. (2026). Graphene Oxide-Based Membranes for Wastewater Treatment: Current Status and Future Outlook. Sustainability & Circularity NOW, 3(continuous publication).
6. Venkata, S. B. (2026). Evidence-Gated Search: Controlling Operational Search Explosion in LLM-Driven Incident Response. Journal of Computer Science and Technology Studies, 8(5), 106-120.
7. Barua, S. (2026). Circular Economy of Source-Separated Organic Waste in New York City. IJST, 5(1).
8. Parasa, M. (2025). SENTINEL-HCM: Federated Temporal Graph Intelligence for Detecting API Abuse, Data Leakage, and Policy Drift in SAP SuccessFactors Cloud Integrations. Journal of Contemporary Science and Technology Management, 1(01), 51-82.
9. U.S. Environmental Protection Agency. (2023). Storm Water Management Model (SWMM) Technical Reference Manual, Version 5.2. EPA Office of Research and Development.
10. Marasani, Y. (2025). Explainable AI Frameworks for Patient-Level Claims Data Analytics. J Artif Intell Mach Learn & Data Sci, 8(1), 3382-3390.
11. Barua, S. (2025). Biochar-Enhanced Filtration Media For Multi-Pollutant Industrial Runoff. Journal of Data Analysis and Critical Management, 1(04), 95-102.
12. Parasa, M. (2024). Architecting predictive workforce intelligence: A machine learning framework for attrition forecasting in SAP Success Factors. Global Scientific and Academic Research Journal of Multidisciplinary Studies, 3(12), 212–221. GSARJMS. https://doi.org/10.5281/zenodo.17587702
13. MARASANI, Y. (2024). Enterprise Readiness for Generative AI: The Critical Role of Data Engineering. Frontiers in Computer Science and Artificial Intelligence, 3(2), 59-71.
14. Mohammadi, S. A. (2023). The Role of Holistic Lifestyle Interventions (Mindfulness, Nutrition, Sleep Optimization and Traditional Therapies) in Improving Stress-Related Disorders and Quality of Life: A Systematic Review. INTERNATIONAL JOURNAL OF APPLIED PHARMACEUTICAL SCIENCES AND RESEARCH, 8(02), 42-54.