Journal of Contemporary Science and Technology Management

Volume 02, Issue 02, December 2026

Integrating CIRx and Machine Learning for Predictive Stormwater and Wastewater Infrastructure Management

Santanu Kumar Dulai
Independent Researcher, New York, USA

Abstract

Municipal stormwater and wastewater utilities manage vast, largely buried asset networks whose deterioration is difficult to observe directly and expensive to inspect exhaustively, forcing asset managers to prioritize inspection and rehabilitation spending under significant uncertainty. This paper presents an integrated framework combining the Condition-Index Rx (CIRx) methodology — a composite, prescriptive condition-scoring approach that translates structural, hydraulic, consequence-of-failure, and operational data into a single actionable rehabilitation priority score — with machine learning (ML) models for predictive deterioration forecasting and anomaly detection. We describe a five-layer data engineering architecture spanning sensing and ingestion, feature and asset-twin storage, ML predictive modeling, CIRx scoring and prioritization, and decision and work-order action, and we detail how CCTV/PACP inspection data, SCADA and flow-sensor telemetry, GIS asset registries, and maintenance history are fused to compute both current condition and forecast risk. A comparative table summarizes the composition and weighting of CIRx components, and a second table compares candidate predictive model classes across precision, recall, and typical forecasting lead time. Two figures illustrate the proposed architecture and an illustrative set of ML-predicted deterioration trajectories across common asset cohorts (cast iron, vitrified clay, reinforced concrete, and PVC/HDPE pipe), showing how predicted trajectories intersect a defined CIRx intervention threshold to trigger rehabilitation prescriptions before failure occurs. We further discuss data governance, model validation against limited failure-event data, and integration with capital planning processes, concluding with open research directions including cross-utility data sharing and network-level cascading-failure modeling.

Keywords: CIRx; condition index; stormwater infrastructure; wastewater infrastructure; machine learning; predictive maintenance; asset management; data engineering; deterioration modeling; GIS.

1. Introduction

Stormwater and wastewater collection systems represent some of the largest and least visible capital assets that municipalities own. Pipe networks, often decades old and constructed from a mix of materials including cast iron, vitrified clay, reinforced concrete, and more recent

PVC or HDPE, deteriorate gradually and largely out of sight, with failure frequently manifesting only as a sinkhole, a sanitary sewer overflow, or a flooded intersection. Utilities have historically relied on periodic closed-circuit television (CCTV) inspection, structural condition coding such as the NASSCO Pipeline Assessment Certification Program (PACP), and reactive maintenance driven by complaint history to manage these assets. While these practices provide valuable condition snapshots, they are inherently limited: exhaustive inspection of an entire network is costly and slow, and condition data alone does not directly answer the question a capital planner actually needs answered, namely which assets should be rehabilitated first, and when.

The Condition-Index Rx (CIRx) methodology addresses part of this gap by combining multiple weighted inputs — structural condition, hydraulic performance, consequence of failure, and operational history — into a single composite score that functions less like a passive index and more like a clinical prescription ("Rx") for infrastructure intervention: a specific, prioritized, actionable recommendation rather than a raw numerical rating alone. Machine learning extends CIRx further by adding a predictive dimension: rather than scoring only the current, observed condition of an asset, ML models can forecast how that condition, and the associated hydraulic and consequence risk, is likely to evolve over a defined planning horizon, allowing CIRx scores to reflect anticipated future risk rather than past inspection findings alone.

This paper makes three contributions. First, it presents an integrated CIRx-plus-ML framework and a five-layer data engineering architecture that supports it, detailing how heterogeneous stormwater and wastewater data sources are fused into a coherent asset-twin representation. Second, it provides a structured comparison, in tabular form, of the CIRx component structure and of candidate predictive model classes suited to different aspects of deterioration and anomaly forecasting. Third, it illustrates the framework through two figures, including an illustrative set of predicted deterioration trajectories across common pipe material cohorts, showing how the integrated system triggers rehabilitation prescriptions ahead of failure.

2. Background

2.1 Condition Assessment and Index-Based Prioritization

Structural condition assessment of buried pipe infrastructure has traditionally relied on CCTV inspection coded against standardized defect taxonomies such as PACP, which assigns structural, operational and maintenance, and overall condition grades to each inspected segment. While valuable, PACP-style condition grades reflect only observed structural defects

at the time of inspection and do not, on their own, incorporate hydraulic performance, the consequence of a given asset's failure, or any forward-looking estimate of deterioration, all of which are necessary for defensible capital prioritization. Composite condition-index approaches address this by combining structural grades with additional weighted factors, an approach on which the CIRx methodology builds by further formalizing the output as an explicit rehabilitation prescription rather than a numerical score alone.

2.2 Machine Learning in Pipe Deterioration Modeling

A growing body of applied research has explored machine learning for wastewater and stormwater pipe deterioration prediction, including tree-based ensemble methods, nearest-neighbor approaches, and neural network models trained on historical inspection and pipe attribute data to predict future condition grades or remaining useful life. These approaches generally outperform simple age-based or linear deterioration curves by incorporating a wider range of explanatory variables, including pipe material, diameter, burial depth, soil characteristics, and prior maintenance history, though they remain constrained by the relative scarcity of confirmed failure events in most utility datasets relative to the size of the asset base.

2.3 Data Engineering Challenges Specific to This Domain

Stormwater and wastewater utilities typically maintain data across multiple, loosely integrated systems: a GIS asset registry, a CCTV/PACP inspection database, a SCADA or telemetry platform for flow and level monitoring, a computerized maintenance management system (CMMS) for work orders, and, increasingly, rainfall and hydraulic model outputs from separate modeling software. Integrating these sources into a coherent, queryable representation suitable for both CIRx scoring and ML model training is a substantial data engineering undertaking in its own right, and is a central focus of the architecture proposed in this paper.

3. The CIRx Framework

CIRx synthesizes five weighted components into a single composite score for each asset segment: structural condition, hydraulic performance, consequence of failure, predicted deterioration, and operational history. Table 1 summarizes each component, its representative data sources, and a typical relative weighting drawn from common asset-management practice, which utilities may adjust to reflect local risk tolerance and regulatory priorities.

Journal of Contemporary Science and Technology Management
Volume 02, Issue 02, December 2026

CIRx Component	Description	Example Data Source	Typical Weight
Structural Condition (S)	PACP-style structural defect severity and density from CCTV/LiDAR inspection	CCTV inspection, PACP coding, LiDAR scans	35%
Hydraulic Performance (H)	Capacity utilization, surcharge frequency, flow/level sensor anomalies	SCADA, flow meters, hydraulic model outputs	20%
Consequence of Failure (C)	Population served, proximity to critical infrastructure, environmental sensitivity	GIS asset registry, land-use data, criticality maps	25%
Predicted Deterioration (P)	ML-forecast probability of failure within defined horizon	Deterioration model output (Layer 3)	15%
Operational History (O)	Prior work orders, maintenance frequency, complaint history	CMMS/work-order system	5%

Table 1. Composite structure of the CIRx (Condition-Index Rx) score.

The distinguishing feature of CIRx relative to a conventional weighted condition index is the explicit inclusion of a predicted deterioration component (P) alongside observed structural condition (S), and the framing of the output not merely as a numeric priority ranking but as a specific rehabilitation prescription — for example, "cured-in-place pipe lining within 18 months" or "targeted point repair within 6 months" — derived from a decision table that maps CIRx score ranges and dominant contributing factors to standard rehabilitation methods. This prescriptive framing is intended to make the output directly actionable by capital planning and operations staff, rather than requiring a separate translation step from score to recommended action.

4. Data Engineering Architecture

Figure 1 presents the proposed five-layer architecture supporting the integrated CIRx-ML framework. Data flows upward from field sensing and inspection through a unified feature and asset-twin store, into ML predictive models, into CIRx composite scoring, and finally into a decision and work-order action layer that interfaces with the utility's capital planning and maintenance systems.

Journal of Contemporary Science and Technology Management
Volume 02, Issue 02, December 2026

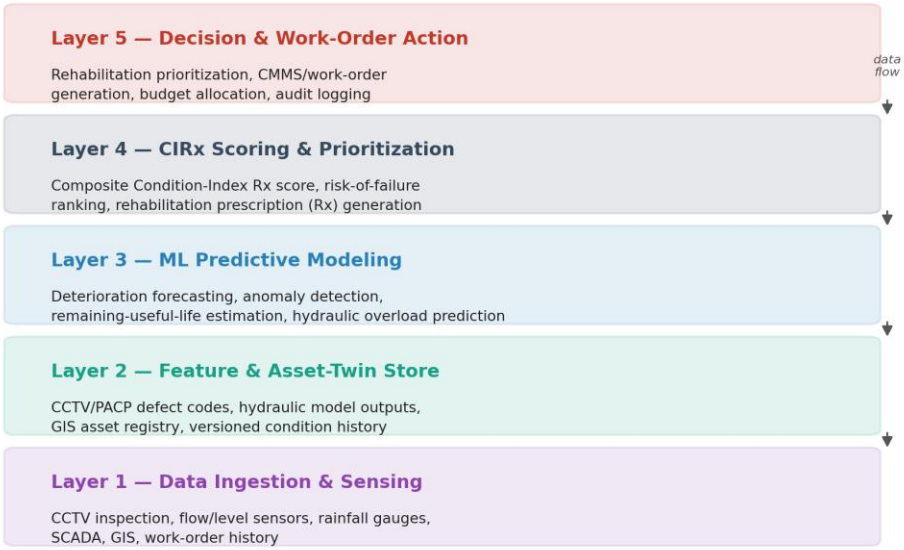

Figure 1. Five-layer data engineering and ML architecture underlying the CIRx framework

4.1 Data Ingestion and Sensing

The ingestion layer aggregates CCTV inspection footage and PACP defect coding, flow and level sensor telemetry, rainfall gauge data, SCADA alarms, and historical work-order records. Given the episodic nature of CCTV inspection, ingestion pipelines must handle both continuous streaming telemetry from sensors and periodic, batch-oriented inspection data on a common ingestion framework, reconciling differing update frequencies without treating stale inspection data as equivalent in confidence to fresh sensor readings.

4.2 Feature and Asset-Twin Store

The feature and asset-twin store maintains a versioned, spatially referenced representation of each pipe segment and associated structure, linking GIS geometry to condition history, hydraulic model outputs, and derived features such as rolling defect density, surcharge frequency, and time since last inspection. Maintaining this store as a versioned asset twin, rather than a simple relational table of current values, allows both CIRx scoring and ML model training to reference the state of an asset as it existed at any prior point in time, which is essential for validating deterioration models against historical outcomes.

4.3 Governance and Data Quality

Because CIRx scores and ML predictions directly inform capital spending decisions, the architecture emphasizes data quality controls at ingestion, including automated validation of PACP coding against defined schema rules, sensor drift detection for flow and level meters, and explicit confidence scoring for inspection data based on inspection age and method, so that downstream scoring and prediction can appropriately discount stale or lower-confidence inputs rather than treating all historical data as equally reliable.

5. Machine Learning for Predictive Deterioration and Anomaly Detection

5.1 Deterioration Forecasting

Deterioration forecasting models predict the probability that a given pipe segment will reach a defined failure or critical-condition threshold within a specified planning horizon, using features drawn from pipe attributes, environmental context, and condition history. Gradient-boosted tree ensembles are commonly used for this task given their strong performance on structured, tabular asset data and their relative interpretability compared to deep learning alternatives, which matters in a regulatory and capital-planning context where model outputs must be explainable to non-technical stakeholders.

5.2 Remaining Useful Life and Survival Modeling

Survival analysis techniques, including random forest survival models, are well suited to remaining-useful-life estimation because they naturally accommodate censored data — the common situation in which most inspected pipe segments have not yet failed at the time of model training — rather than requiring a binary failed/not-failed label that discards information about segments still in service.

5.3 Sensor-Based Anomaly Detection

For hydraulic performance monitoring, unsupervised anomaly detection methods such as isolation forests and autoencoders, alongside recurrent neural network architectures such as long short-term memory (LSTM) networks, are applied to flow and level sensor streams to detect early signs of blockage, infiltration and inflow, or capacity exceedance, often providing lead time measured in hours to days ahead of a visible surcharge or overflow event.

Journal of Contemporary Science and Technology Management
Volume 02, Issue 02, December 2026

5.4 Network-Level Risk Propagation

Graph neural network approaches, operating over the pipe network topology maintained in the asset-twin store, support network-level risk propagation analysis, identifying segments whose failure would be likely to cause cascading downstream capacity or structural impacts, which is difficult to capture through segment-level modeling alone. Table 2 compares these model classes across illustrative precision, recall, and typical forecasting lead time, based on patterns commonly reported in applied deterioration-modeling literature.

Model Type	**Task**	**Precision**	**Recall**	**Typical Lead Time Gained**
Gradient-boosted trees	Pipe-segment deterioration classification	0.81	0.77	2–4 years
Recurrent neural network (LSTM)	Sensor-based hydraulic anomaly forecasting	0.78	0.83	Hours to days
Random forest survival model	Remaining-useful-life estimation	0.75	0.72	3–6 years
Graph neural network	Network-level cascading failure risk propagation	0.79	0.74	1–3 years
Isolation forest / autoencoder	Unsupervised sensor anomaly detection	0.70	0.85	Hours to days

Table 2. Comparison of candidate predictive model classes for CIRx-integrated deterioration and anomaly forecasting.

6. Illustrative Scenario: Predicted Deterioration Trajectories

Figure 2 illustrates predicted deterioration trajectories for four representative asset cohorts — cast iron/ductile iron, vitrified clay pipe, reinforced concrete, and PVC/HDPE — over a twenty-year forecast horizon, expressed as an ML-predicted probability of failure. A horizontal intervention threshold marks the point at which the CIRx framework is configured to trigger a rehabilitation prescription. In this illustrative scenario, the cast iron cohort crosses the intervention threshold earliest, consistent with its generally shorter expected service life relative to the other materials modeled, prompting an earlier CIRx-generated prescription for targeted rehabilitation, while the PVC/HDPE cohort remains below the threshold for the longest portion of the forecast horizon, reflecting its comparatively longer expected service life.

Journal of Contemporary Science and Technology Management
Volume 02, Issue 02, December 2026

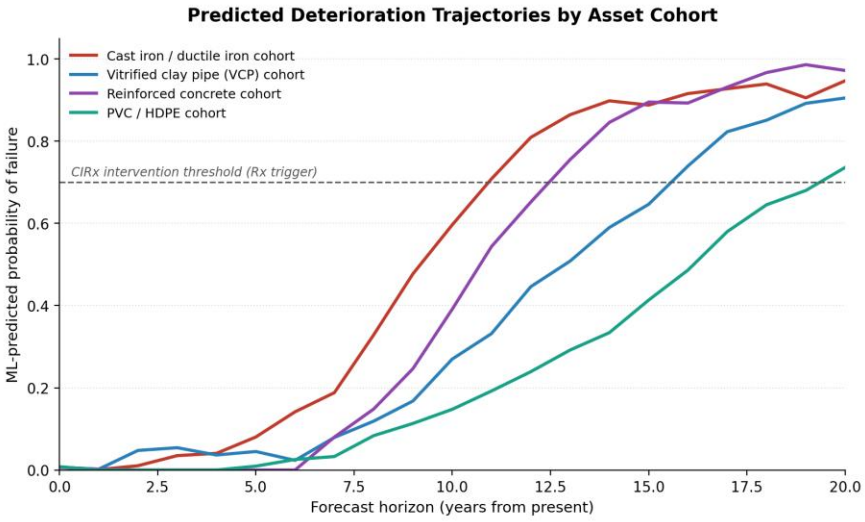

Figure 2. Illustrative ML-predicted deterioration trajectories used to trigger CIRx rehabilitation prescriptions

This scenario demonstrates the core value proposition of integrating ML forecasting with CIRx: rather than waiting for a scheduled inspection to reveal that a segment has already reached a poor structural condition grade, the framework flags the segment for prescriptive action once its predicted trajectory is forecast to cross the intervention threshold within the utility's defined planning horizon, shifting capital planning from a reactive, inspection-triggered process toward a proactive, forecast-triggered process.

To make this concrete, consider a single cast-iron segment currently rated as structurally sound on its most recent CCTV inspection, carrying a low structural condition component (S) but located within a combined sewer catchment with a documented history of surcharge events, giving it an elevated hydraulic performance component (H) and, owing to its location beneath a major arterial road, a high consequence-of-failure component (C). Under a condition-index approach relying on structural grade alone, this segment might not surface near the top of a rehabilitation priority list, since its observed structural condition remains acceptable. Under the integrated CIRx-ML framework, however, the predicted deterioration component (P) — informed by the gradient-boosted deterioration model's forecast that this segment's cohort trajectory is projected to cross the intervention threshold within roughly eleven years, as illustrated for the cast-iron cohort in Figure 2 — combines with the already-elevated H and C components to produce a composite CIRx score sufficient to trigger a prescription for proactive lining well ahead of any structural defect becoming visible on inspection. This example illustrates why the predicted deterioration component is not merely a convenience but a materially different input than structural condition alone: it allows the composite score to reflect forward-looking risk for exactly the segments where consequence and hydraulic stress are already elevated, rather than waiting for structural evidence to accumulate before acting.

7. Integration with Capital Planning

For the framework to influence real capital outcomes, CIRx-generated prescriptions must integrate with the utility's existing capital improvement planning and CMMS work-order processes rather than operating as a standalone analytical exercise. In practice, this means the decision and action layer exports ranked prescriptions, together with supporting CIRx component scores and model confidence, into the planning tools capital engineers already use, preserving human review and budget-constrained prioritization rather than assuming unlimited rehabilitation capacity. Because rehabilitation budgets are always constrained relative to the full set of flagged assets, the framework also supports scenario analysis, allowing planners to compare the network-wide risk reduction achieved under different funding levels and prioritization rules before committing a capital program.

8. Discussion and Open Challenges

8.1 Validating Models Against Sparse Failure Data

A persistent challenge in this domain is that confirmed failure events are rare relative to the overall size of most utility asset bases, which constrains the amount of positive-labeled training data available for supervised deterioration models and motivates continued use of survival-analysis and unsupervised techniques that make more efficient use of censored and unlabeled data.

8.2 Cross-Utility Data Sharing

Individual utilities, particularly smaller ones, often lack sufficient historical failure data to train robust deterioration models independently, suggesting a role for anonymized, cross-utility data pooling or shared regional deterioration models, though this remains constrained by data governance concerns and inconsistent condition-coding practices across jurisdictions.

8.3 Explainability for Regulatory and Public Accountability

Because CIRx prescriptions inform public capital spending decisions, model outputs feeding into the composite score must remain explainable to capital planners, elected officials, and the public, favoring model classes and score decompositions that can be clearly attributed to specific structural, hydraulic, or consequence factors rather than opaque end-to-end predictions.

8.4 Climate Adaptation and Changing Hydraulic Loading

Climate-driven changes in rainfall intensity and frequency are shifting the hydraulic loading assumptions embedded in legacy stormwater design, meaning that deterioration and capacity models trained purely on historical data risk understating future risk; incorporating forward-looking climate projections into the hydraulic performance component of CIRx is an important area for continued refinement.

9. Conclusion

This paper has presented an integrated framework combining the CIRx composite condition-scoring methodology with machine learning-based predictive modeling to support proactive stormwater and wastewater infrastructure management. By formalizing a five-layer data engineering architecture that unifies inspection data, sensor telemetry, GIS asset records, and maintenance history into a coherent asset-twin representation, and by extending CIRx scoring with ML-forecast deterioration and anomaly signals, utilities can shift capital planning from a reactive, inspection-triggered process toward a proactive, forecast-triggered one that generates specific, actionable rehabilitation prescriptions ahead of failure. Realizing this shift in practice requires sustained attention to data quality, model explainability, and integration with existing capital planning workflows, and future work should prioritize cross-utility data-sharing mechanisms and the incorporation of forward-looking climate projections into deterioration and hydraulic performance forecasting.

References

Venkata, S. B. (2026, January). CIRx: Autonomous Healing for Build and Deploy. In The International Conference on Artificial Intelligence and Smart Environment (pp. 299-313). Cham: Springer Nature Switzerland.

Vladeanu, G., & Matthews, J. (2019). Wastewater pipe condition rating model using multicriteria decision analysis. Journal of Water Resources Planning and Management, 145(12), 04019058.

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.

NASSCO. (2021). Pipeline Assessment and Certification Program (PACP) reference manual. National Association of Sewer Service Companies.

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.

Marasani, Y. (2025). Explainable AI Frameworks for Patient-Level Claims Data Analytics. J Artif Intell Mach Learn & Data Sci, 8(1), 3382-3390. U.S. Environmental Protection Agency. (2022). Clean Watersheds Needs Survey (CWNS) report.

Barua, S. (2025). Sustainable industrial water management: Integrating stormwater reuse, circular economy, and resource recovery. British Journal of Environmental Studies, 5(3), 08-22.

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..

Kipf, T. N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. International Conference on Learning Representations.

Venkata, S. B. (2026, January). Runbook Mesh: MCP-Orchestrated Terraform and Ansible Co-Execution on Azure. In 2026 Second International Conference on Intelligent Systems for Communication, IoT and Security (ICISCoIS) (pp. 1-7). IEEE.

Barua, S. (2024). Reactive Soil Mixes for Enhanced PFAS Adsorption in Stormwater Infiltration Basins: Mechanisms and Field Assessment. *SAMRIDDHI: A Journal of Physical Sciences, Engineering and Technology*, *16*(01), 60-66.

Grieves, M., & Vickers, J. (2017). Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems. In Transdisciplinary Perspectives on Complex Systems (pp. 85–113). Springer.

Barua, S. (2024). REAL-TIME IO T-ENABLED CONTROL OF STORMWATER ASSETS: REDUCING RUNOFF PEAKS AND POLLUTANT LOADS. Multidisciplinary Innovations & Research Analysis, 5(4), 100-120.