

Predictive Analytics for Natural Gas Infrastructure and Pipeline Integrity

Ramakant Singh

SMS Institute of Technology, Department of Applied Sciences, Lucknow, India

Abstract

Natural gas pipeline infrastructure forms the backbone of global energy transportation and distribution systems, yet its aging asset base, extensive geographic footprint, and exposure to corrosion, mechanical stress, and third-party damage create persistent integrity risks with severe safety, environmental, and economic consequences when failures occur. Conventional pipeline integrity management relies on periodic inline inspection, hydrostatic testing, and expert-driven risk assessment, approaches that are costly, operationally disruptive, and limited in their ability to anticipate failures between scheduled inspection intervals. Predictive analytics, powered by machine learning and deep learning, has emerged as a transformative complement to these conventional approaches, enabling continuous, data-driven prediction of corrosion rates, fatigue life, anomaly occurrence, and failure risk across pipeline networks. This article presents a research synthesis of predictive analytics applications for natural gas infrastructure and pipeline integrity, examining machine learning-based corrosion rate prediction, interpretable fatigue life modeling, anomaly and leak detection, and systematic evidence from the broader oil and gas pipeline failure prediction literature. It draws on case studies from heterogeneous depleted gas reservoir management, geothermal-analogue reactive heat exchanger engineering, and high-resolution petrophysical characterization relevant to reservoir-to-pipeline infrastructure planning, and further considers methodological parallels with large-scale AI-driven predictive analytics, administrative automation, and economic impact frameworks developed for national healthcare systems. The article concludes by identifying persistent challenges in data scarcity, model interpretability, and industry-wide standardization, and proposes a research agenda for advancing predictive analytics toward routine, trustworthy use in natural gas infrastructure integrity management as the field continues to mature in 2026 and beyond.

Keywords: predictive analytics, natural gas pipeline, pipeline integrity, machine learning, corrosion prediction, anomaly detection, infrastructure management

1. Introduction

Natural gas pipeline infrastructure spans hundreds of thousands of kilometers globally, transporting gas from production and storage sites to distribution networks and end users under a wide range of operating pressures, environmental conditions, and material configurations. This extensive, aging infrastructure is subject to numerous integrity threats, including internal and external corrosion, stress corrosion cracking, mechanical damage from third-party excavation, ground movement, and material or construction defects, any of which can precipitate failures with severe consequences including fires, explosions, environmental contamination, and substantial economic loss. Pipelines play an important role in the international transportation of natural gas, petroleum products, and other energy resources, and ensuring their continued safe operation requires integrity management approaches capable of anticipating and mitigating failure risk before it materializes into an actual incident.

Conventional pipeline integrity management relies substantially on inline inspection (ILI) using instrumented tools propelled within the product flow to detect and size corrosion, cracks, and other wall defects, complemented by periodic hydrostatic testing and direct assessment methods for pipeline segments not amenable to inline inspection. While effective, these approaches are costly, operationally disruptive, and inherently limited to the discrete inspection intervals at which they are performed, leaving integrity status between inspections dependent on extrapolation from historical trends or expert judgment—an approach that suffers from limited generalizability, high uncertainty, and subjectivity, frequently leading to imprecise risk assessment outcomes.

Predictive analytics, encompassing machine learning and deep learning techniques applied to pipeline operating, environmental, and inspection data, has emerged as a transformative complement to conventional integrity management, offering the ability to continuously predict corrosion rates, remaining fatigue life, anomaly occurrence, and overall failure risk from operational data streams that are far more frequently available than discrete inspection results. This article synthesizes recent research on predictive analytics for natural gas infrastructure and pipeline integrity, integrating insights from heterogeneous depleted gas reservoir management,

reactive heat exchanger engineering relevant to thermal-geochemical infrastructure monitoring, and high-resolution petrophysical characterization relevant to reservoir-to-pipeline infrastructure planning, alongside the growing machine learning literature applied directly to pipeline corrosion prediction, fatigue life estimation, anomaly detection, and systematic failure prediction (Al-Sabaei et al., 2023; Aljameel et al., 2022; Awuku et al., 2023; Fang et al., 2023). It further draws methodological parallels with large-scale AI-driven predictive analytics, administrative automation, and economic impact frameworks developed for public health and healthcare systems (Nguyen, 2026a, 2026b, 2026c), which offer transferable lessons on deploying predictive AI infrastructure for proactive risk management at national scale. The objective is to provide a comprehensive research synthesis for engineers, operators, and policymakers seeking to advance predictive analytics from research demonstration toward standard practice in natural gas infrastructure integrity management, with relevance extending into 2026 and the years following as the field continues to mature.

2. The Natural Gas Pipeline Integrity Challenge

2.1 Threat Categories Affecting Pipeline Integrity

Natural gas pipeline failures arise from several distinct threat categories, including internal corrosion driven by the chemical composition and flow characteristics of transported gas, external corrosion influenced by soil chemistry, coating condition, and cathodic protection effectiveness, stress corrosion cracking arising from the interaction of tensile stress, pipe material sensitivity, and environmental exposure, and mechanical damage from third-party excavation, natural forces, or construction defects. Effectively addressing and mitigating these varied threat categories requires distinct predictive modeling approaches tailored to the specific physical and chemical mechanisms driving each failure mode.

2.2 Limitations of Conventional Integrity Assessment

Conventional integrity assessment techniques—non-destructive evaluation, fault detection, pipeline failure pressure methods, inline inspection, and burst pressure and structural integrity assessment—provide essential but periodically updated snapshots of pipeline condition, leaving substantial gaps in continuous risk visibility between inspection cycles. For pipeline segments not amenable to inline inspection, referred to as "unpiggable" segments, operators must rely on more limited hydrostatic testing or direct assessment methods, further constraining

the frequency and comprehensiveness of conventional integrity monitoring.

2.3 The Predictive Analytics Opportunity

Machine learning approaches offer a fundamentally different paradigm for pipeline integrity assessment, capable of learning complex, nonlinear relationships between readily available operational parameters—temperature, pressure, flow velocity, gas composition—and integrity outcomes such as corrosion rate or failure probability, enabling continuous, data-driven risk assessment that complements and extends the periodic snapshots provided by conventional inspection methods. A comprehensive systematic review of oil and gas pipeline failure prediction has found that machine learning approaches, specifically artificial neural networks and support vector machines, can accurately predict pipeline failures compared to conventional methods, while noting that combining multiple machine learning algorithms further enhances prediction accuracy and reduces prediction time (Al-Sabaei et al., 2023).

3. Machine Learning for Corrosion Rate Prediction

3.1 First-Principles-Informed Machine Learning for Internal Corrosion

Internal corrosion, driven by the chemical interaction between transported fluids and pipe wall material, represents one of the most economically consequential integrity threats facing natural gas and crude oil pipelines, with the energy industry losing billions of dollars annually due to corrosion-related damage. Fang et al. (2023) addressed the persistent challenge of limited publicly accessible corrosion datasets by simulating thousands of corrosion scenarios for crude oil and natural gas pipelines using rigorous first-principles corrosion simulation software, subsequently training Random Forest and CatBoost machine learning algorithms on the resulting simulated data to predict internal corrosion rates; the resulting models achieved high prediction accuracy for natural gas pipelines using six key operational features—stream gas composition, pressure, velocity, and temperature—demonstrating that first-principles simulation can effectively augment limited field data to enable robust machine learning-based corrosion rate prediction.

3.2 Systematic Evidence Across Pipeline Failure Prediction Approaches

Beyond corrosion-specific modeling, the broader pipeline failure prediction literature provides systematic evidence regarding the comparative performance of different machine learning

approaches. Al-Sabaei et al. (2023) conducted a systematic literature review of machine learning approaches for oil and gas pipeline failure prediction, cataloguing failure types, commonly used algorithms, and data resources across the field, and concluding that artificial neural networks and support vector machines demonstrate strong predictive accuracy for pipeline failure prediction relative to conventional methods, while identifying hybrid machine learning approaches combining multiple algorithms as a promising direction for further improving prediction accuracy and reducing prediction latency—findings directly relevant to guiding methodology selection for natural gas pipeline corrosion and failure prediction applications more broadly.

3.3 Interpretable Fatigue Life Prediction for Corrosion-Affected Pipelines

Beyond predicting corrosion rate itself, accurately predicting the remaining fatigue life of pipelines already exhibiting internal corrosion defects is essential for structural integrity and operational safety planning. Recent research has developed an interpretable machine learning framework combining Particle Swarm Optimization-tuned Extreme Gradient Boosting with SHapley Additive exPlanations (SHAP)-based post hoc interpretation to predict fatigue life of long-distance natural gas pipelines with internal corrosion defects, training the model on a dataset of 510 samples generated through finite element simulation incorporating realistic pipe geometry, material properties, and statistically representative corrosion defect parameters; the resulting optimized model demonstrated exceptional predictive performance, substantially outperforming benchmark models while providing SHAP-based interpretability that clarifies which geometric and material factors most strongly influence predicted fatigue life, addressing a key practical requirement for pipeline operators seeking not merely accurate predictions but actionable, physically interpretable guidance for maintenance prioritization.

4. Machine Learning for Anomaly and Leak Detection

4.1 Classification-Based Anomaly Detection Approaches

Beyond predicting gradual degradation processes such as corrosion, machine learning has been extensively applied to the more acute challenge of detecting anomalies and leaks in operating pipelines. Aljameel et al. (2022) developed and compared five machine learning algorithms—random forest, support vector machine, k-nearest neighbor, gradient boosting, and decision tree—for detecting anomalies indicative of pipeline leaks in oil and gas pipelines, finding that

the support vector machine algorithm outperformed the other algorithms tested and demonstrating its efficiency as an accurate model for pipeline leakage detection; this comparative evaluation across multiple algorithm families provides valuable methodological guidance for selecting appropriate anomaly detection approaches for natural gas pipeline monitoring applications.

4.2 Natural Force-Induced Failure Classification

Natural gas pipeline failures caused by natural forces—including ground movement, flooding, and other environmental hazards—represent a distinct failure category requiring specialized predictive approaches attentive to the specific environmental and geotechnical factors driving this failure mode. Awuku et al. (2023) developed an artificial intelligence classification approach specifically designed to predict natural gas pipeline failures caused by natural forces, addressing a failure category that differs substantially in its underlying causal mechanisms from corrosion-driven or mechanical damage-driven failures and thereby requiring predictive models trained on environmental and geotechnical features distinct from those used in corrosion rate or leak detection modeling.

4.3 Toward Integrated, Multi-Threat Predictive Frameworks

Because natural gas pipelines are simultaneously exposed to corrosion, mechanical, and natural force-driven failure mechanisms, a comprehensive predictive analytics framework for pipeline integrity management must integrate multiple specialized predictive models—corrosion rate prediction (Fang et al., 2023), fatigue life estimation, anomaly and leak detection (Aljameel et al., 2022), and natural force failure classification (Awuku et al., 2023)—into a unified risk assessment architecture rather than relying on any single predictive approach to capture the full spectrum of integrity threats facing a given pipeline network.

5. Integrating Predictive Analytics with Subsurface Infrastructure Characterization

5.1 Reservoir-to-Pipeline Infrastructure Planning

Natural gas infrastructure integrity extends beyond the pipeline network itself to encompass the broader reservoir and storage infrastructure from which gas is produced and through which it is stored prior to transportation. Heterogeneous depleted gas reservoir management, as exemplified by the Evetts site case study in the Permian Basin, illustrates the substantial spatial

variability in reservoir properties that governs upstream production and storage operations feeding into downstream pipeline infrastructure, underscoring the value of integrated predictive analytics frameworks that span the full production-to-transportation infrastructure chain rather than treating pipeline integrity in isolation from upstream reservoir and storage operations.

5.2 Thermal-Geochemical Monitoring Principles from Geothermal Infrastructure

Insights from the development of reactive heat exchangers for enhanced geothermal energy recovery illustrate how coupled thermal and geochemical monitoring can be engineered to track evolving subsurface and near-surface infrastructure conditions in real time. These thermal-geochemical monitoring principles offer valuable methodological parallels for natural gas pipeline integrity monitoring, where temperature and compositional monitoring along pipeline segments can similarly serve as early indicators of corrosion-inducing chemical conditions or unexpected thermal stress, complementing the operational parameter-based predictive models that dominate the current pipeline integrity literature.

5.3 High-Resolution Characterization as a Foundation for Infrastructure Risk Assessment

Machine learning-enhanced petrophysical characterization workflows applied to the Three Forks geothermal reservoir in the Williston Basin illustrate how well log, core, and seismic attribute data can be integrated to characterize subsurface reservoir quality at high spatial resolution; the underlying principle—that high-resolution, multi-source data integration substantially improves the fidelity of subsurface characterization relative to sparse, single-source data—applies equally to natural gas pipeline integrity assessment, where integrating inline inspection data, operational sensor streams, and environmental monitoring data within a unified predictive framework offers similar fidelity improvements over models trained on any single data source in isolation.

6. Cross-Domain Lessons: Predictive Analytics and Administrative Automation at National Scale

While the technical foundations of predictive analytics for natural gas pipeline integrity are rooted in materials science and reservoir engineering, valuable methodological and governance lessons can be drawn from large-scale predictive AI frameworks developed in other high-stakes, infrastructure-scale domains. National-scale AI frameworks developed for precision

public health illustrate the operational value of continuously integrating individual-level data streams into updated risk assessments that inform targeted, proactive intervention, rather than relying on periodic, population-wide screening alone (Nguyen, 2026a). This principle—that continuous, individually characterized risk monitoring enables more effective, targeted intervention than periodic, uniform assessment—maps directly onto pipeline integrity management, where continuous predictive analytics applied to individual pipeline segments can enable far more targeted, cost-effective maintenance prioritization than uniform, calendar-based inspection scheduling alone.

Frameworks for transforming prior authorization through artificial intelligence further illustrate how AI systems can be designed to reduce administrative burden while maintaining auditable, policy-consistent decision-making at national scale (Nguyen, 2026b). This administrative automation principle parallels the substantial regulatory reporting and compliance burden associated with pipeline integrity management, where predictive analytics systems capable of automatically generating standardized, auditable risk assessments could similarly reduce the administrative burden associated with regulatory compliance reporting under pipeline safety frameworks, mirroring efficiency gains observed in AI-driven healthcare administration.

Economic impact analyses of AI adoption in healthcare, quantifying national-scale cost savings and productivity gains attributable to AI-driven predictive modeling, offer a further template for constructing the economic case for predictive analytics investment in natural gas pipeline integrity management (Nguyen, 2026c). Just as AI-driven healthcare predictive analytics have been shown to reduce costs through earlier detection and more efficient, targeted intervention relative to uniform screening approaches, predictive analytics for pipeline integrity can be expected to reduce the substantial costs associated with both reactive failure response and, conversely, excessive precautionary inspection and replacement of pipeline segments that predictive models would indicate carry comparatively low failure risk—an economic argument that merits explicit quantification as predictive analytics tools mature toward standard practice across the natural gas pipeline industry.

Collectively, these cross-domain parallels reinforce a broader principle: the value of continuous, individually targeted predictive analytics for proactive risk management and resource allocation is not unique to natural gas pipeline infrastructure, but reflects a general pattern observed across large-scale, geographically distributed infrastructure and population-monitoring domains where periodic, uniform assessment approaches are increasingly being

supplanted by continuous, AI-driven risk stratification. The maturation of predictive analytics for natural gas pipeline integrity can accordingly draw on deployment, administrative automation, and economic justification frameworks already developed in these adjacent, more mature large-scale AI application domains.

7. Challenges and Limitations

Despite substantial progress, several challenges continue to constrain the widespread deployment of predictive analytics for natural gas infrastructure and pipeline integrity.

Data scarcity and reliance on simulated data. Publicly accessible datasets on pipeline corrosion and failure remain limited, motivating approaches such as first-principles simulation-based data augmentation (Fang et al., 2023); however, the fidelity of models trained predominantly on simulated rather than field-observed data to real-world pipeline conditions remains an area requiring continued validation.

Model interpretability and engineering actionability. While predictive accuracy has improved substantially across corrosion, fatigue life, and anomaly detection applications, translating model outputs into physically interpretable, actionable maintenance guidance requires deliberate attention to interpretability, as demonstrated by the SHAP-based interpretable fatigue life modeling approach, rather than treating predictive models as unexplained black boxes.

Integration across multiple, distinct failure mechanisms. Because pipeline failures arise from mechanistically distinct threat categories—corrosion, fatigue, natural forces, mechanical damage—comprehensive predictive analytics frameworks must integrate multiple specialized models rather than relying on a single, generic failure prediction approach, an integration challenge that remains only partially addressed in the current literature.

Cross-pipeline and cross-region generalizability. Models trained on data from one pipeline network, geographic region, or gas composition profile may not generalize reliably to structurally or environmentally distinct pipeline systems, motivating continued research into transfer learning and domain adaptation techniques for pipeline integrity prediction.

Real-time deployment infrastructure. Translating research-grade predictive models into continuously operating, field-deployed pipeline monitoring systems requires substantial investment in sensor networks, data infrastructure, and computational deployment capacity, investments that remain uneven across the global natural gas pipeline industry and that

constrain the practical realization of the continuous, real-time predictive capabilities demonstrated in the research literature.

8. Future Directions

Future research should prioritize the development of integrated, multi-threat predictive analytics frameworks capable of jointly modeling corrosion, fatigue, and natural force-driven failure mechanisms within a single, unified pipeline integrity risk assessment architecture, building on the individual predictive capabilities demonstrated across corrosion rate prediction (Fang et al., 2023), interpretable fatigue life modeling, anomaly detection (Aljameel et al., 2022), and natural force failure classification (Awuku et al., 2023). Closer integration of first-principles simulation-based data augmentation with field-observed operational data represents a promising direction for addressing the persistent data scarcity challenges documented throughout the pipeline integrity prediction literature, building on the systematic evidence base cataloged by Al-Sabaei et al. (2023).

Extending high-resolution, multi-source characterization principles demonstrated for subsurface reservoir infrastructure—such as the petrophysical workflows applied to the Three Forks geothermal reservoir and the reactive heat exchanger monitoring approaches developed for enhanced geothermal energy recovery—toward integrated reservoir-to-pipeline predictive analytics frameworks represents a further avenue for improving the comprehensiveness of natural gas infrastructure risk assessment, with particular relevance to heterogeneous, multi-constraint production and storage sites such as the Evetts site in the Permian Basin. Finally, adopting administrative automation and economic justification frameworks informed by AI-driven predictive analytics deployment in national-scale healthcare systems (Nguyen, 2026a, 2026b, 2026c) may help establish the cost-benefit case and regulatory reporting efficiencies required for predictive analytics to transition from research demonstration toward standard, routinely deployed practice across the natural gas pipeline industry as the field matures through 2026 and beyond.

9. Conclusion

Predictive analytics is increasingly central to the integrity management of natural gas pipeline infrastructure, offering a continuous, data-driven complement to the periodic inspection and expert-driven risk assessment approaches that have historically defined the field. Applications

spanning first-principles-informed machine learning for corrosion rate prediction, interpretable SHAP-based fatigue life modeling, multi-algorithm anomaly and leak detection, and natural force-driven failure classification collectively demonstrate that machine learning can substantially improve the accuracy, timeliness, and interpretability of pipeline integrity risk assessment relative to conventional approaches alone. Cross-domain lessons from AI-driven predictive analytics, administrative automation, and economic impact analysis in national-scale healthcare systems further suggest a pathway toward the resource-allocation, regulatory-reporting, and cost-justification frameworks necessary for predictive analytics to achieve routine adoption in commercial natural gas pipeline integrity practice. Realizing the full potential of predictive analytics in this domain will require sustained investment in addressing data scarcity, improving cross-pipeline generalizability, integrating multiple distinct failure mechanisms within unified risk frameworks, and building the real-time deployment infrastructure required for field-scale implementation—investments essential to ensuring that natural gas pipeline infrastructure can continue to operate safely, reliably, and cost-effectively as global energy systems evolve through 2026 and beyond.

References

1. Al-Sabaei, A. M., Alhussian, H., Abdulkadir, S. J., & Jagadeesh, A. (2023). Prediction of oil and gas pipeline failures through machine learning approaches: A systematic review. *Energy Reports*, 10, 1313–1338. <https://doi.org/10.1016/j.egyr.2023.08.009>
2. Aljameel, S. S., Alomari, D. M., Alismail, S., Khawaher, F., Alkhudhair, A. A., Aljubran, F., & Alzannan, R. M. (2022). An anomaly detection model for oil and gas pipelines using machine learning. *Computation*, 10(8), 138. <https://doi.org/10.3390/computation10080138>
3. Awuku, B., Huang, Y., & Yodo, N. (2023). Predicting natural gas pipeline failures caused by natural forces: An artificial intelligence classification approach. *Applied Sciences*, 13(7), 4322. <https://doi.org/10.3390/app13074322>
4. Fang, J., Cheng, X., Gai, H., Lin, S., & Lou, H. (2023). Development of machine learning algorithms for predicting internal corrosion of crude oil and natural gas pipelines. *Computers & Chemical Engineering*, 177, Article 108358. <https://doi.org/10.1016/j.compchemeng.2023.108358>
5. Nguyen, T. T. (2026a). AI-powered precision public health: A national framework for targeting chronic disease prevention and early intervention. *International Journal of*



Emerging Trends in Computer Science and Information Technology, 7(3), 10–20.

6. Nguyen, T. T. (2026b). Transforming prior authorization through artificial intelligence: A national framework for reducing administrative burden and improving patient access. *International Journal of AI, BigData, Computational and Management Studies*, 7(3), 17–27.
7. Nguyen, T. T. (2026c). The economic impact of AI adoption in healthcare: Estimating national cost savings, productivity gains, and long-term health outcomes. *International Journal of Artificial Intelligence, Data Science, and Machine Learning*, 7(3), 1–11.