



Journal of Advanced Research in Technology and Innovation Management

Journal homepage:
<https://karyailham.com.my/index.php/jartim/index>
ISSN: 2811-4744



Developing a Machine Learning Approach for intelligent Pipeline Pigging Maintenance in Oil and Gas

Ahmed Hossain Abdulkhair Ben Salem¹, Asmaa Mahfoud Hezam Al-Hakimi^{1,*}

¹ Faculty of Information Sciences and Engineering Management and Science University 40100 Shah Alam, Selangor, Malaysia

ARTICLE INFO

Article history:

Received 21 August 2026

Received in revised form 30 August 2026

Accepted 30 September 2026

Available online 12 October 2026

Keywords:

Pipeline pigging; machine learning; digital twin; reinforcement learning; predictive maintenance

ABSTRACT

This article examines governance challenges and digital transparency reforms in Algeria's hydrocarbon sector from an institutional perspective. Given the sector's central role in the national economy, weak governance, limited transparency, and persistent corruption have constrained effective resource management and economic diversification. Drawing on a descriptive-analytical approach, the study analyzes institutional reports, legal frameworks, governance indicators, and recent digital initiatives to assess the evolution of governance practices. The findings reveal a significant gap between Algeria and successful resource-rich countries in terms of accountability, regulatory quality, and control of corruption. While recent reforms such as digital compliance platforms, whistleblowing mechanisms aligned with international standards, and national anti-corruption strategies signal a gradual shift toward improved transparency, their impact remains limited by structural and institutional constraints. The study argues that digital transformation can enhance governance only when embedded within broader institutional reforms. Strengthening transparency, enforcement, and oversight is essential for improving the investment climate, reducing rent-seeking behavior, and promoting sustainable management of hydrocarbon resources in Algeria.

1. Introduction

Pipeline systems are a fundamental component of the global energy sector and represent the primary means of transporting oil and gas over long distances. These systems form vast and complex networks that connect production fields to refineries, processing plants, storage terminals, and final consumers. Like the circulatory system in a living organism, pipelines continuously transport crude oil, natural gas, and refined petroleum products, ensuring the uninterrupted flow of energy that supports industrial activity, economic development, and daily life. From an economic perspective,

* Corresponding author.

E-mail address: asmaa@msu.edu.my

pipelines offer the most cost-effective and efficient method for transporting large volumes of hydrocarbons compared to road, rail, or marine transportation. This efficiency enables the exploitation of remote oil and gas reserves and contributes to the stability of regional and global energy markets [1]. Strategically, pipeline infrastructure plays a critical role in national energy security by providing a reliable and continuous supply of energy that is less exposed to geopolitical instability, supply chain disruptions, or extreme weather conditions. In addition, pipelines are recognized as one of the safest land-based transportation methods for oil and gas, with most systems installed underground, reducing exposure to human populations and minimizing environmental disturbance [2].

Despite these advantages, the operation and maintenance of oil and gas pipelines face numerous technical, operational, and environmental challenges. Pipeline integrity is continuously threatened by various degradation mechanisms that develop over time under harsh operating conditions. One of the most significant threats is corrosion, which gradually weakens pipeline walls and increases the likelihood of leaks or ruptures. Internal corrosion is commonly caused by the presence of water, carbon dioxide (CO₂), hydrogen sulfide (H₂S), and microbial activity within the transported fluids, while external corrosion is influenced by soil properties, moisture content, coating degradation, and stray electrical currents. These corrosion processes may result in uniform wall thinning or localized pitting corrosion, both of which are difficult to detect and predict accurately using conventional inspection approaches [3]. These deposits can accumulate on the inner pipeline walls, reducing the effective diameter, increasing pressure drop, restricting flow, and in severe cases causing complete blockages. The situation is further complicated by the aging nature of pipeline infrastructure, as many pipelines currently in operation have exceeded their original design life and were constructed using older materials and standards.

To manage pipeline integrity and maintain operational efficiency, the oil and gas industry relies extensively on pigging operations. Pigging involves inserting a device known as a pig, or Pipeline Inspection Gauge, into the pipeline and propelling it through the system using the pressure and flow of the transported product. The term “pig” is believed to originate from the squealing sound produced by early pipeline cleaning tools as they moved through pipelines. Over time, pigging technology has evolved significantly, leading to the development of various pig types designed for specific operational purposes [4]. Utility and cleaning pigs, such as foam pigs and mandrel pigs equipped with brushes or blades, are used to remove debris, wax, and other deposits. More advanced Intelligent Pigs, also referred to as In-Line Inspection (ILI) tools, are equipped with high-precision sensors that enable non-destructive testing of pipeline conditions. Common ILI technologies include Magnetic Flux Leakage (MFL) tools for detecting metal loss due to corrosion, Ultrasonic Testing (UT) tools for measuring wall thickness, and caliper tools for identifying geometric deformations. Although these inspection tools provide valuable condition data, traditional pigging practices remain largely reactive and periodic. Pigging schedules are typically based on fixed time intervals rather than the actual condition of the pipeline, and ILI inspections provide only a snapshot of pipeline health at a specific moment [5].

Recent advances in Machine Learning (ML) and data-driven technologies have opened new opportunities to overcome the limitations of traditional pigging and maintenance practices. Several studies have demonstrated the effectiveness of ML in enhancing pipeline monitoring and predictive maintenance. The integrating deep learning techniques into intelligent pigging systems significantly improves defect detection and enables early prediction of pipeline failures, allowing maintenance activities to be planned more effectively. Predictive maintenance models can optimize pigging schedules, reduce unplanned downtime, and lower operational risks by shifting maintenance from time-based to condition-based strategies. The value of ML by demonstrating that machine learning

models can accurately predict corrosion, leaks, and other integrity threats, leading to improved maintenance decision-making and enhanced pipeline safety [6].

In parallel with ML advancements, digital twin technology has emerged as a powerful tool for pipeline integrity management. A digital twin is a virtual representation of a physical pipeline system that continuously updates using real-time sensor data and operational information. The digital twins enable real-time monitoring and simulation of pipeline behavior under varying operating conditions, allowing potential failures to be identified before they occur. emphasized that predictive simulations can be used to optimize pig movement, cleaning efficiency, and operational safety. In addition, a theory-guided machine learning framework capable of accurately predicting pitting corrosion depth, highlighting the importance of combining physical knowledge with data-driven models [7].

This project focuses specifically on crude oil pipelines, which are highly vulnerable to wax deposition and flow assurance challenges. Wax accumulation can lead to increased pressure drop, reduced flow efficiency, and unplanned shutdowns, making it an ideal use case for predictive pigging. The proposed project adopts a simulation-based approach that integrates machine learning, predictive scheduling, autonomous pig control, and a digital twin environment. By anticipating wax buildup, optimizing pigging schedules, and improving pig movement control, the system aims to enhance operational efficiency and reduce maintenance risks. This focused scope allows for realistic modeling of flow dynamics, pig–pipeline interaction, and wax deposition behavior while maintaining manageable simulation complexity [8].

2. Literature Review

Artificial Intelligence is a general area in computer science that is concerned with the development of systems capable of performing tasks that demand human-like intelligence, which would involve reasoning, learning, perception, and decision making. Machine Learning is a large subdivision of AI that entails the development of algorithms that find patterns in data and make predictions independently of any prior programmability to do so. Some useful ML methods in pipeline maintenance include supervised learning as to predict corrosion or wax deposition, unsupervised learning as to detect anomalies, and reinforcement learning as to optimize pigging schedules and control policies. Such techniques are useful in transforming raw data collected in the pipeline into useful insights [9].

Distributed sensors and smart pigging tools are IoT devices that collect valuable information such as pressure, temperature, vibration, acoustic signature, and magnetic flux leakage. This sensor network is like a digital nervous system, which gives the continuous data flow that is necessary in intelligent analysis. Predictive and autonomous maintenance systems could not be visible and effective due to the lack of IoT [10]. Predictive maintenance is a major milestone in contrast to the conventional ways of maintenance. Reactive maintenance just reacts to the failures that have already happened, whereas preventive maintenance is performed according to strict schedules no matter the real state of the pipeline. Conversely, predictive maintenance involves the use of real time and past information to forecast failures prior to their occurrence. In the case of pipeline operations, predictive maintenance is useful in minimizing downtime, blockage, and the occurrence of severe failure.

The convergence of AI and IoT is transforming the operations of pipelines into smart and interconnected systems. IoT offers high-frequency data, and AI is an analytical engine, in which anomalies are detected, future risks predicted, and operational decisions are enhanced. With this combination, autonomous or semi-autonomous control is possible where the pipelines can constantly self-check their status and take the necessary corrective measures. Industry 4.0 principles

also enhance this integration with cutting-edge analytics, digital twins and data-driven processes. The importance of intelligent pipeline maintenance is manifested in perhaps several ways. Pipeline blockages, emergency repairs and ineffective pigging can lead to serious financial losses economically. Intelligent predictive systems are also useful in mitigating these risks by allowing their early identification and minimizing operational interruptions [11].

Nevertheless, the oil and gas industry continues to struggle with such challenges as unpredictable wax deposition, lack of real-time visibility within pipelines, inefficient pigging schedules, and disjointed Machine Learning applications, which can solve problems, but do not offer any real-time functions.

2.1 Theoretical Foundations and Relevant Models

This section explains the major theories and models that will be used to design and implement the intelligent pipeline maintenance system. The knowledge of these theories is beneficial to the successful application of predictive maintenance, sensor analytics, and machine learning methods. This will guarantee integrity of the pipeline, optimisation of pigging schedule and minimisation of operational risks.

- **Condition Monitoring Theory**

The condition monitoring theory is a theory proposed by the American Psychologist. Condition Monitoring Theory: The condition monitoring theory is a theory advanced by an American Psychologist. Predictive maintenance is based on the Condition Monitoring (CM) theory. It pays attention to the constant monitoring and evaluation of such operational parameters as temperature, pressure, vibration, and flow rate. The point of this is to detect the onset of wear or abnormal behaviour to prevent failure occurrence. In pipeline systems, CM combines sensor networks and data acquisition tools to perform real time monitoring to check the internal conditions. Identifying anomalies and lack of normal operation, CM can be used to support proactive maintenance intervention strategies, reduce the likelihood of serious failures, and support data-driven decision-making [12].

- **Supervised Learning Theory**

A machine learning method that involves the use of labelled datasets to train models is known as supervised learning. All inputs are connected to a well-known output. This model trains to relate input features with outcomes making it predict future events with accuracy. In the project, controlled learning forecasts the rates of wax deposition, the dangerous pipeline sections, and the probability of the pigging requirements. The training data is historical operational data that has temperature readings, pressure variation, flow rate variation, and the readings of wax. Continuous variables such as the thickness of wax are predicted using regression techniques whereas segments will be classified as either Safe, At Risk, or Critical. This contributes to targeted maintenance measures [13].

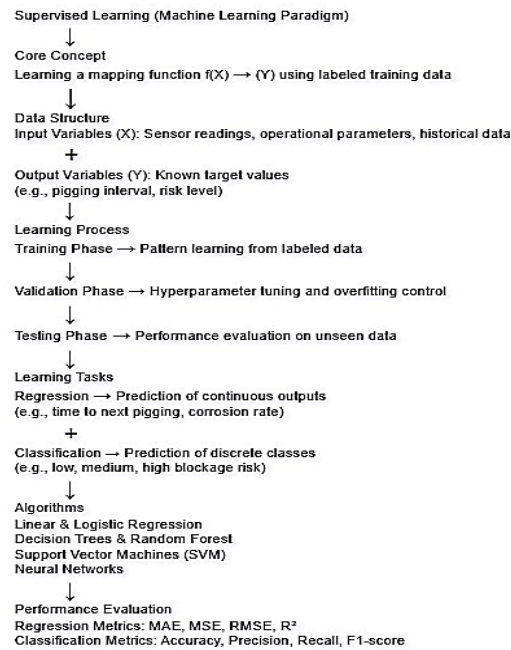


Fig. 1. Supervised Learning Theory in machine learning

Figure 1. illustrates the theoretical arrangement of supervised learning in the machine learning system. It initiates with the main idea to learn a useful relationship between the features of the input and the known output labels relying on labeled datasets. The data model is a collection of input variables denoting quantifiable system parameters and output variables denoting target predictions.

The learning process is structured into 3 key steps namely training, validation and testing. Training The model acquires patterns using historical, labeled data. The validation stage is aimed at maximizing the model parameters to avoid overfitting and the testing stage is to test the predictive accuracy of the model on unseen data. The chart also differentiates the supervised learning tasks into regression and classification which will be based on whether the target of prediction is continuous or categorical. Figure 1. also displays the popular algorithms of supervised learning and correlates them with suitable performance assessment measures [14].

- Flow Assurance Principles

Flow assurance principles are vital for ensuring the smooth and efficient transport of hydrocarbons through pipelines. These principles explain the interactions among fluid properties, temperature, pressure, and flow behaviour. They are especially important for understanding the creation of wax, hydrates, and other flow-related blockages. Using flow assurance theory lets engineers model the rate of wax deposition, predict potential blockage sites, and optimize pipeline operating conditions. These principles provide the engineering foundation for combining predictive models with real-time sensor data to maintain pipeline efficiency and safety [15].

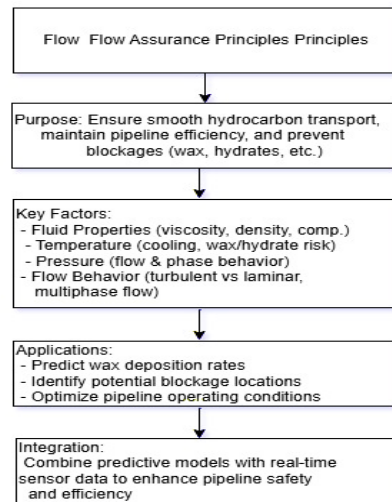


Fig. 2. Flow assurance principles

As illustrated in figure 2. Flow Assurance Principles ensure hydrocarbons flow efficiently through pipelines by managing fluid properties, temperature, pressure, and flow behaviour. They help predict and prevent blockages like wax and hydrates, optimize operating conditions, and integrate predictive models with real-time sensor data for safe and efficient pipeline operation.

- Predictive Analytics in Pipeline Maintenance

Predictive analytics transforms raw operational and sensor data into actionable insights. It uses statistical analysis, data mining, and machine learning algorithms to identify trends, detect anomalies, and forecast future events. In pipeline maintenance, predictive analytics supports data-driven decision-making by predicting the optimal times for pigging operations, estimating the remaining useful life of pipeline segments, and identifying sections at risk of corrosion or blockage. By integrating real-time sensor data with historical operational records, predictive analytics enables proactive, efficient, and cost-effective maintenance strategies [16].

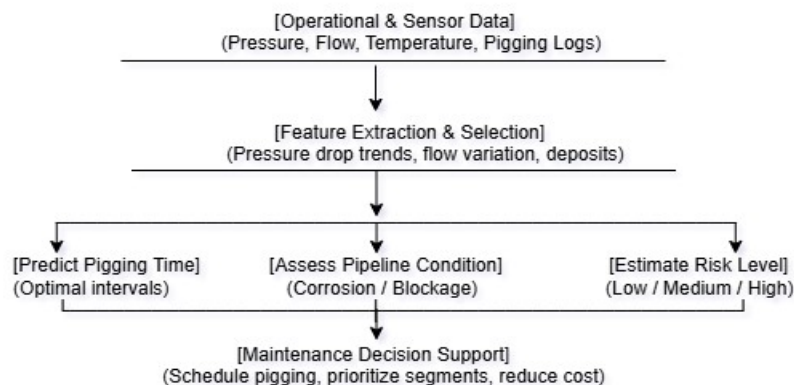


Fig. 3. Data-driven pipeline maintenance using predictive analytics

Figure 3. presents a data-driven approach to pipeline maintenance using predictive analytics. Operational and sensor data are first analyzed to extract relevant features related to pipeline behaviour. These features are then used to predict optimal pigging intervals, assess pipeline condition, and estimate maintenance risk levels. The combined outputs support informed maintenance decisions, enabling proactive pigging scheduling, improved pipeline safety, and reduced operational costs [17].

2.2 Relevant Algorithms and Methodologies

A wide range of algorithms and methodologies have been applied to address challenges in pipeline pigging, predictive maintenance, and autonomous control, both in research and industry. In predictive maintenance and anomaly detection, traditional and advanced machine learning (ML) algorithms have been extensively employed. Random Forests and Support Vector Machines (SVM) to forecast pipeline failures and optimize pigging schedules, leveraging historical inspection data and operational parameters such as pressure, temperature, and flow rate. Automatic machine learning models for maximum pitting corrosion depth prediction. A robust ML technique for wax deposition prediction. ML models for pipeline integrity assessment, highlighting the role of predictive analytics in proactive pipeline management. Deep learning approaches, including CNNs and RNNs, have been applied for signal processing, acoustic-based inspection, and multiphase flow behaviour modelling allowing the detection of corrosion patterns, flow anomalies, and wax accumulation [18].

In autonomous pig control, reinforcement learning (RL) algorithms, including Q-Learning, Deep Q-Networks (DQN), and policy gradient methods, have been explored to optimize pig movement and minimize operational risks, such as stalling or excessive energy consumption. Reward functions are designed to balance efficiency, safety, and pipeline integrity, allowing intelligent agents to adapt to complex and dynamic flow conditions within simulated or partially real environments.

Additionally, predictive scheduling methodologies combining probabilistic modelling, optimize pigging intervals based on operational uncertainty and risk assessment. These approaches provide uncertainty-aware predictions, addressing a gap in traditional deterministic scheduling. Hybrid approaches that integrate ML predictions, digital twin simulations, and hardware sensor data have also been proposed to overcome individual limitations, creating a unified, proactive pipeline maintenance system. Collectively, these methodologies form a comprehensive foundation for intelligent, predictive, and autonomous pipeline maintenance, addressing gaps identified in both academic research and industry practice [10].

2.3 Implementation Approaches

Previous studies and industrial systems have implemented a variety of approaches to pipeline monitoring, predictive maintenance, and pigging operations. Hardware-based implementations dominate traditional pipeline inspection practices. Systems such as smart pigs, caliper pigs, foam pigs, and bi-directional pigs deployed by companies like SGS, Halliburton, Baker Hughes, ROSEN Group, PETRONAS Carigali, and Uzma Engineering are highly effective for detecting corrosion, wax deposits, and mechanical obstructions. The primary strength of these approaches is high accuracy and reliability in real-time measurements, with direct feedback from physical pipelines. However, their weaknesses include high deployment cost, operational disruption, and reactive nature, as inspections typically detect problems only after they occur. Long pipelines require repeated deployments, which is labour-intensive and time-consuming, limiting scalability [11].

In contrast, simulation-based and digital twin implementations integrate computational models such as CFD, FEA, multiphase flow simulations, and physics-informed neural networks to predict pig behavior, wax deposition, and pipeline stresses. These approaches provide a safe, cost-effective, and flexible platform to test scenarios, plan maintenance schedules, and evaluate new pig designs without risking physical assets. The main advantage is their ability to simulate complex operational conditions and train intelligent agents for autonomous control. However, the effectiveness of these implementations is limited by insufficient validation against real operational data, model simplifications, and challenges in integrating live sensor data from pipelines.

Machine learning-based implementations have been widely applied for predictive maintenance, failure detection, and optimized pigging schedules. Models such as Random Forests, Gradient Boosting, SVM, Neural Networks, CNNs, and RNNs use historical and real-time operational data to provide proactive insights. Their strength lies in forecasting potential failures and improving maintenance planning, while also handling large datasets and extracting complex patterns. Weaknesses include dependence on high-quality, representative datasets, reduced accuracy when trained on limited or simulated data, and limited real-time integration with hardware and simulation platforms in current industrial practice. Hybrid implementation approaches that combine hardware data, machine learning predictions, and digital twin simulations (Baker Hughes, Halliburton, Shell, Saudi Aramco) attempt to leverage the strengths of each approach while mitigating individual weaknesses. These integrated frameworks allow proactive, data-driven, and validated pipeline monitoring and maintenance, but their adoption is still limited due to high complexity, resource requirements, and integration challenges across different platforms and technologies. Such a framework can address operational gaps, reduce costs, improve safety, and provide real-time, autonomous decision-making in pipeline pigging operations.

2.4 State of the Art in the Field

- Recent Advances

The field of pipeline monitoring and pigging has witnessed significant advancements over the past few years, driven by both academic research and industrial innovation. Hardware-based inspection systems have evolved beyond traditional caliper pigs to incorporate smart pigs, bi-directional pigs, and intelligent caliper foam pigs capable of real-time data collection and advanced detection. Industrial leaders such as SGS, Halliburton, Baker Hughes, ROSEN Group, PETRONAS Carigali, and Uzma Engineering have deployed these technologies, integrating embedded sensors and telemetry systems to capture pressure, flow, corrosion, and wax deposition data during pipeline operation.

In parallel, digital twin and simulation-based approaches have become increasingly sophisticated. Researchers and industries have developed physics-informed simulation models combining multiphase flow dynamics, structural mechanics, and pig-pipeline interactions. These models enable scenario analysis, predictive maintenance scheduling, and testing of operational strategies without risking pipeline integrity. Advanced methods such as Ensemble Kalman Filters (EnKF) and Physics-Informed Neural Networks (PINNs) allow assimilation of real-time sensor data into simulations, enhancing accuracy and reliability. Machine learning and AI-driven solutions have also gained traction. Techniques such as Random Forests, Gradient Boosting, SVM, Neural Networks, CNNs, and RNNs are used for predictive maintenance, anomaly detection, wax deposition prediction, and failure forecasting. Industrial implementations in Shell and Saudi Aramco demonstrate the practical impact of AI in improving operational safety, efficiency, and predictive scheduling for pigging operations. Reinforcement learning has recently been applied for autonomous pig control, enabling intelligent agents to optimize pig movement in real-time under dynamic pipeline conditions.

- Trends and Future Directions

Emerging technologies such as quantum computing offer potential for solving highly complex optimization problems in predictive maintenance and scheduling, particularly for large pipeline networks with multiple interacting variables. Additionally, advanced data analytics, cloud-based monitoring, and IoT integration are shaping the field, providing scalable platforms for continuous monitoring and predictive analysis. The combination of physics-informed ML, digital twins, and real-time hardware feedback is projected to dominate future research, aiming to create proactive,

autonomous, and highly reliable pipeline maintenance systems that can significantly enhance safety, reduce costs, and optimize operational efficiency across global oil and gas infrastructure.

2.6 Critical Review of Similar Products or Systems

Several projects in both academia and industry have targeted pipeline inspection, pigging optimization, and predictive maintenance. In academic research, systems such as predictive maintenance frameworks using machine learning algorithms have been developed to forecast pipeline failures, wax deposition, and corrosion depth. Simulation-based approaches using digital twins, computational fluid dynamics (CFD), finite element analysis (FEA), and multiphase flow modelling have been proposed to replicate pipeline–pig interactions and evaluate operational scenarios safely. In the industry, companies such as Shell, Saudi Aramco, Halliburton, Baker Hughes, SGS, ROSEN Group, PETRONAS Carigali, and Uzma Engineering have deployed smart pigging solutions that combine hardware-based inspection systems with limited predictive analytics. For example, Shell’s digital twin and machine learning solutions allow pipeline operators to predict maintenance needs, while Saudi Aramco leverages ML for anomaly detection in pigging operations. Other companies like SGS and Halliburton focus on real-time sensor-based monitoring with advanced caliper and foam pigs for detecting deposits, corrosion, and mechanical defects.

The proposed project significantly extends current systems by integrating three critical components: predictive scheduling using uncertainty-aware machine learning, a physics-informed digital twin for pig-flow-structure simulation, and autonomous pig control via reinforcement learning. Unlike existing academic models, which often rely on limited data sets or simulations, the proposed system aims to incorporate real-time hardware data from pipeline operations. Compared to industrial systems, which are largely reactive and hardware-centric, this project is proactive, adaptive, and intelligent, capable of optimizing pigging schedules, predicting anomalies with quantified uncertainty, and autonomously controlling pig operations in real-time. Additionally, the integration of physics-informed models with ML and RL agents provides a novel framework that addresses gaps in accuracy, adaptability, and safety present in current solutions. Comparison between Similar Systems and Proposed System is illustrated in Figure 4.

Feature / Aspect	Existing Academic Systems	Existing Industrial Systems	Proposed System
Data Source	Historical datasets, limited sensor data	Real-time sensors, hardware feedback	Historical + real-time hardware sensor data integrated with simulation
Predictive Capability	ML prediction of failures and wax deposition	Basic scheduling, some anomaly detection	Uncertainty-aware ML for predictive pigging and maintenance planning
Simulation	CFD, FEA, multiphase flow (limited validation)	Rarely used for predictive planning	Physics-informed digital twin validated with real pipeline data
Automation / Control	Limited, often simulation-only	Manual or semi-automated pig operations	Reinforcement learning–based autonomous pig control with adaptive decision-making
Integration	Partial (ML or simulation only)	Hardware-centric with minimal analytics	Fully integrated: ML + digital twin + autonomous control

Fig. 4. Matrix comparison among current systems and proposed system

This critical review demonstrates that while existing solutions provide important functionality in pipeline monitoring and maintenance, there remain gaps in predictive accuracy, integration, and autonomous control. The proposed system addresses these limitations by combining hardware-based real-time data, advanced predictive algorithms, digital twin simulations, and adaptive control, offering a comprehensive, state-of-the-art pipeline maintenance framework.

2.6 Summary of Findings

The literature and industry review reveal several common themes in pipeline pigging, monitoring, and maintenance. First, hardware-based inspection systems, such as smart pigs, caliper pigs, and foam pigs, remain the backbone of industrial practice, providing reliable real-time measurements of pipeline conditions, including corrosion, wax deposition, and mechanical obstructions. Companies like SGS, Halliburton, Baker Hughes, ROSEN Group, PETRONAS Carigali, and Uzma Engineering have successfully implemented these systems, demonstrating high measurement accuracy and operational reliability.

Second, simulation-based and digital twin approaches are increasingly applied to replicate complex pipeline–pig–fluid interactions. These methods allow safe testing of scenarios, planning of maintenance schedules, and validation of predictive algorithms without risk to physical pipelines. Additionally, machine learning techniques including Random Forests, Gradient Boosting, SVMs, Neural Networks, CNNs, and RNNs have been widely adopted for predictive maintenance, wax deposition forecasting, and anomaly detection. These algorithms demonstrate strong predictive capabilities but are highly dependent on data quality, integration with real-time hardware, and model validation. Existing academic studies often rely on limited or simulated datasets, reducing the applicability of their predictive models in real-world conditions.

2.7 Research Gap

From the review, the key research gaps can be summarized as follows:

1. **Integration Gap:** Current systems rarely combine hardware-based monitoring, predictive machine learning, and digital twin simulations into a unified framework.
2. **Predictive Accuracy Gap:** Most industrial solutions rely on deterministic schedules or simple thresholds, lacking uncertainty-aware prediction and risk-informed scheduling.
3. **Autonomy Gap:** There is a lack of intelligent, autonomous control of pigging operations, which could optimize efficiency, safety, and resource usage in real-time.
4. **Validation Gap:** Many simulation-based and AI-driven methodologies have limited validation against real operational data, reducing their reliability and industrial adoption.
5. **Operational Efficiency Gap:** Hardware deployments are costly and disruptive, and existing systems do not sufficiently minimize downtime or optimize maintenance intervals.

These gaps collectively demonstrate the necessity and relevance of the proposed project. By integrating uncertainty-aware predictive scheduling, a physics-informed digital twin, and reinforcement learning–based autonomous pig control, the project addresses all critical limitations in current research and industrial practice.

3. Methodology

This section involves the research methodology used in the development of an intelligent pipeline pigging maintenance system based on the method of machine learning, digital twin simulation, and

reinforcement learning. This chapter aims at outlining the methodology used to evaluate the constraints of traditional pigging operation and design-predictive and autonomous solutions and test the outcomes of these solutions in a simulated pipeline scenario.

The research will be carried out in six sequential methodological steps. Business knowledge will entail setting the objectives, involve predicting pigging periods, minimizing downtime, and decreasing costs, and determining key performance indicators (KPIs), such as mean absolute error (MAE). During the in-line inspection data understanding part, past pigging operation records, sensor logs, and in-line inspection reports will be reviewed to determine the data quality and missing data. The preparation of data will include cleaning, normalizing, imbalanced event handling and pigging performance relevant engineering. The modelling step will involve the construction and the test of machine learning models, such as random forest, XGBoost and LSTM networks, and cross-validation to guarantee accuracy. Performance will be evaluated through precision, recall, simulation-based ROI analysis, and deployment will be through creation of a maintenance schedule recommender, and development of documentation to be handed to the operations team.

The use of a simulation-based digital twin environment will be used in the research due to the safety hazards and the associated high operational costs of physical pipelines. Such an environment will enable predictive and autonomous models to be tested in a controlled manner and give information on the system performance in these conditions of a real pipeline without undermining safety or overly high costs. Performance measures will be implemented through the relevant metrics, and comparison outcomes will be made with the traditional means of performing the same measure to be able to show the advantages of the suggested approach.

3.1 Dataset Description and Validity

This research will utilize a comprehensive dataset combining historical pigging records, in-line inspection (ILI) reports, and real-time pipeline sensor measurements, including pressure, temperature, flow rate, and operational logs.

3.1.1 Typical attributes within the dataset include:

- Pipeline specifications: age, diameter, material, length
- Operational parameters: pressure, flow rate, temperature
- Pig characteristics: type, size, foam or caliper, run speed
- Maintenance indicators: pigging interval, detected corrosion, wax thickness, anomalies
- Sensor readings: real-time pressure, flow fluctuations, temperature trends

The dataset is considered valid and reliable because it represents actual pipeline operating conditions and provides quantitative measurements directly related to pigging performance and maintenance decisions. It enables objective analysis, supports predictive modelling, and allows calibration and testing of the digital twin simulation environment.

3.1.2 Usage of the dataset:

- Training and validating machine learning models for predictive pigging scheduling.
- Calibrating and testing the physics-informed digital twin to simulate pig–pipeline interactions.
- Developing reinforcement learning agents for autonomous pig control.

3.1.3 Data limitations and handling:

Potential limitations include missing values, data imbalance, and heterogeneous formats between ILI reports and sensor outputs. These will be addressed through data cleaning, normalization, feature engineering, and probabilistic methods, ensuring the dataset remains robust and suitable for predictive modelling and simulation.

3.1.3 Dataset validity justification

The dataset is considered valid and suitable for this research because:

- It represents real pipeline operating conditions
- It provides objective, measurable indicators directly related to pigging performance and maintenance decisions
- Similar datasets have been successfully used in previous studies to develop predictive maintenance models and digital twin simulations

The selected dataset supports reliable machine learning model development and realistic simulation-based validation.

3.1.4 Data preparation and analysis method

In the sub-subsection, the processing and analysing of the gathered pipeline datasets will be described to aid machine learning modelling and simulation-based validation of the maintenance of intelligent pigging.

3.1.5 Data preparation

To guarantee accuracy, consistency and appropriateness in machine learning applications, the data obtained on the pipeline will be subjected to a systematic data preparatory procedure before analysis. The initial one is data cleaning, which will be done by identifying partial, inconsistent or noisy data of in-line inspection report and operational sensor log. Missing values: The interpolation of missing values will be done by using suitable statistical imputation methods or interpolations, based on the type of data. The operational parameters that will be subjected to the process of data normalization and scaling will include pressure, flow rate, temperature, and pig speed so that every feature will be fairly represented when training the model. This is a necessary step since in most cases pipeline datasets include variables whose units, as well as value ranges, differ. This will be followed by feature engineering to identify meaningful indicators about maintenance of pigging. Derived features can be the speed deviation of pigs, the level of pressure gradient change, the time elapsed since the last pigging process, and corrosion stage indicators.

3.1.6 Machine learning analysis

Following the preprocessing, the resulting ready data will be utilized to create machine learning models of predictive pigging maintenance and scheduling. The techniques of supervised learning will be used, since the historical data of pipeline includes labelled information concerning maintenance events and intervals of pigging. According to the results of the research conducted on smart pipeline maintenance in the past, the following algorithms should be used:

- a. Random Forest, the complex, nonlinear relationships, and the influential feature identification.
- b. ANNs, to model the dynamic behaviour of a pipeline and sensor interaction.

- c. Support Vector Machines, in the classification task of maintenance risk and pigging necessity.

The data will be split into training and testing groups to have objective evaluation of the models. Standard evaluation metrics to be used to measure model performance will include Mean Absolute Error (MAE) when performing a predictive task and accuracy or precision when performing a classification-based outcome.

3.1.7 Simulation and validation

To confirm the usefulness of the machine learning predictions, a digital twin environment will be created using simulations. This simulation will simulate the conditions under which the pipeline operates such as flow behaviour and pig movement with real operational data being used to calibrate the simulation. The machine learning models will be tested in the simulation to provide predictions of schedules of pigging to determine their feasibility, safety, and effect on operations. The proposed system can be measured in various operating conditions with this method of validation, enabling it to be more reliable and applicable in the industry. Overall, the combination of the data preparation and analysis process includes:

- a. Automatic pre-processing of real pipeline data.
- b. Predictive Modelling based on machine learning.
- c. A digital twin framework of simulation-based validation.

This solution will promote the fact that the suggested intelligent pipeline pigging maintenance system will be data-driven, technically sound, and agreed upon by academic literature as well as industry practices.

3.2. Proposed System Requirements

This section stipulates the needs of the proposed intelligent pipeline pigging system maintenance. The requirements are based on the research objectives, the gaps in the literature studies realized and the limitations which are seen in the current industry solutions to pipeline pigging and integrity management. The proposed system will help service the needs of the pipeline engineers, maintenance planners, and decision-makers through predictive assistance, simulation-based assistance, and intelligent assistance in pigging operations. As opposed to the current systems which primarily aim at inspection or relying on rules to schedule a machine, the proposed system considers machine learning, physics-based simulation, and autonomous decision-making to enhance operational efficiency and safety.

3.2.1 Functional requirements

The offered system will be able to:

1. Gather and catalog historical and operating pipeline data, and in-line inspection (ILI) reports, sensor data, and records of pigging operations.
 - a. Clean, normalize and extract features of preprocess pipeline datasets to make them suitable to be analyzed by machine learning.
 - b. Determine the optimal schedules of pigging with uncertainty-sensitive machine learning models according to the pipeline conditions and the working conditions.
 - c. Measure uncertainty of prediction to help in making risk-informed maintenance decisions.

- d. Run pigging simulations in a physics-informed digital twin simulating pipeline flow behaviour and pig behaviour.
2. Test and train reinforcement learning agents to control autonomous pig movement in the simulation environment.
3. Deliver supportable outputs and visualization to assist in engineering and maintenance planning.

3.2.3 Non-functional requirements

The following non-functional requirements shall be met in the system:

- Accuracy: The predictions should be precise enough to decide on maintenance in engineering.
- Scalability: The system needs to be able to process high amounts of pipeline and sensor data.
- Safety: No physical control of physical pipeline assets should be performed but only in simulation environment during testing and validation.
- Usability: The system output must be comprehensible by the engineers without necessarily having advanced machine learning skills.
- Reliability: The system should be able to perform under different conditions of data quality.
- Data Security: Pipeline operating information should be treated in a secure and confidential manner.

3.2.4 Data requirements

- The system demands actual operational and historical pipeline data, and includes:
- In-line inspection (ILI) statistics (corrosion indicators, anomalies, pig tracking)
- Data on the operation sensors (pressure, temperature, flow rate)
- Past pigging routine and service history.
- Physical and operational properties of pipelines.

The data requirements are compatible with the ones typically employed in past academic research and industrial pipeline integrity system. Overall, the specified system requirements will guarantee that the offered solution will meet the needs of a real industrial application and will offer enhanced machine learning, simulation, and autonomous control solutions. It is based on these requirements that the proposed system design will be formulated as discussed in the next section.

3.2.5 Proposed system design

This section outlines conceptual design of the proposed intelligent pigging maintenance system of the pipeline. The system is built to be a modular and data-driven system and combines predictive analytics, digital twin simulation, and autonomous control to overcome the drawbacks of current pipeline maintenance systems.

3.3 System Overview

The suggested system is based on a layered architecture, which divides data management, analytics, simulation, and decision support functions. The modular design improves flexibility, scalability and maintainability besides maintaining compatibility with the industrial pipeline operations. System Architecture and Main Components. The proposed system will be comprised of the following main components:

3.3.1 Data management layer

This layer does the ingestion of data, storage and preprocesses. It manages ILL reports, sensors and previous pigging data and guarantees the quality and consistency of data prior to analysis.

3.3.2 Machine learning layer

The layer uses predictive pigging maintenance scheduling models. It involves the utilization of past and current data to come up with risk-aware forecasts, including the use of uncertainty estimation to aid in sound decision making.

3.3.3 Digital twin simulation layer

The digital simulates the real-life behaviour of the pipeline, fluid motion and the pig dynamics. It facilitates the testing and validation of predictive models with safety under various operating conditions and cases to counter the gap of integrated simulation environment provided in the literature.

3.3.4 Autonomous control layer

This layer uses reinforcement learning to create an intelligent agent that will optimize the movement of pigs. The agent only communicates with the digital twin environment, which provides safe and controlled training and evaluation.

3.3.5 Decision support interface

This element displays the forecasted findings, simulation findings, and work insights in a readable and understandable format to the engineers and maintenance planners.

3.3.6 System data flow

The pipeline data is initially handled through the data management layer and then it is sent to the machine learning layer where it goes through the predictive analysis. The forecasts are then assessed and verified in the digital twin simulation layer. The autonomous control agent is in this simulated environment to exploit pigging behaviour. Lastly, results of the validated insights are displayed to the users via the decision support interface.

3.3.7 Conformance to research objectives

Objective 1 (Predictive Scheduling): The machine learning layer answered this objective. Objective 2 (Digital Twin Simulation): Was taken care of by the simulation layer. Objective 3 (Autonomous Control): The autonomous control layer addresses this objective. The system design offers a scalable and integrated system design of intelligent pigging maintenance of the pipeline. The approach to data-driven prediction, physics-based simulation, and autonomous control is critical in filling the gap in the literature and industry as the system is safe, reliable, and practically applicable.

3.3.7 System architecture

The package diagram embraces the UML dependency relationships in the form of the stereotype «uses» to illustrate functional relationships among the modules of the system. The packages are

loosely coupled and designed as modular systems since each package relies solely on the services it needs of the other packages. The system architecture is illustrated in Figure 5.

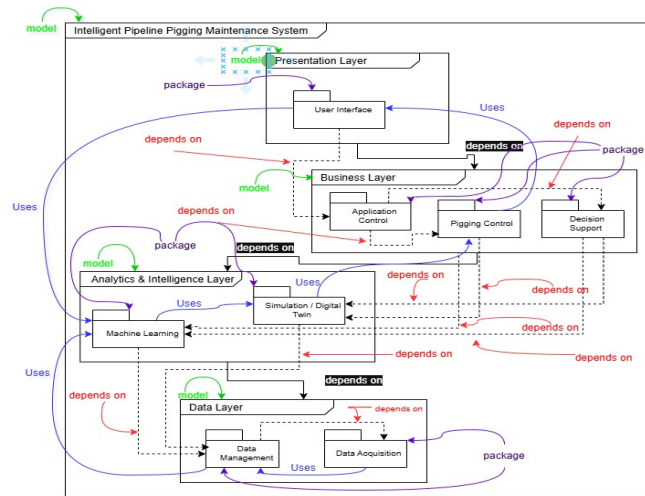


Fig. 5. Package diagram for intelligent pipeline pigging maintenance system

4. Results and Findings

4.1 Machine Learning Predictive Scheduling Results

The predictive scheduling component was developed using a Random Forest (RF) classifier to detect anomalies and assess pipeline risks based on real-time operational data. To ensure robust learning, the model was trained on a combined dataset comprising 111,000 samples, integrating 100,000 physically consistent synthetic data points, 10,000 samples from the AI4I 2020 Predictive Maintenance dataset, and 1,000 samples from the Mississippi State SCADA Pipeline dataset.

4.1.1 Machine learning output:

The final production model (Version 2) comprises 100 decision trees ($n_{\text{estimators}} = 100$, $\text{max_depth} = 15$) and evaluates seven features to classify the pipeline condition into six distinct states: normal, corrosion, leak, friction, stuck_risk, and blockage. The model achieved a high overall accuracy of 98.05% during testing on a holdout set of 22,200 samples.

- ML prediction graph

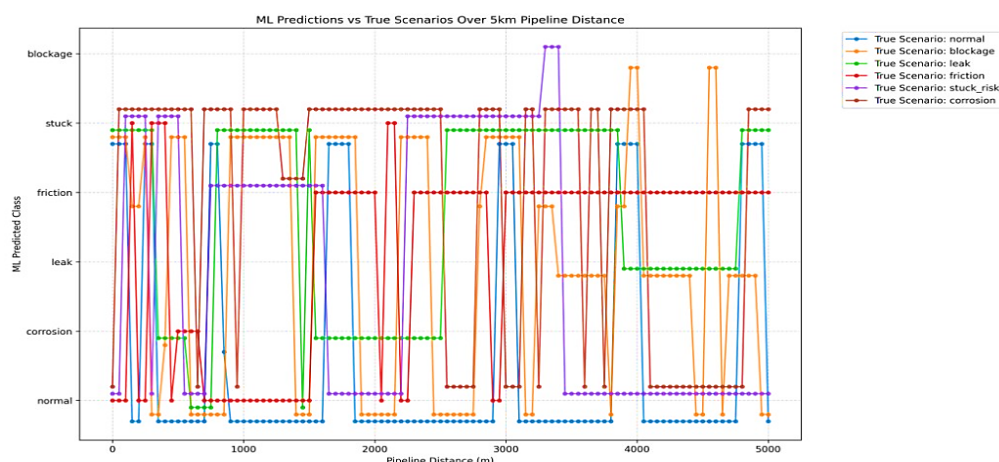


Fig. 6. ML predictions over 5km pipeline

The X-axis represents the physical distance down the pipeline (0 to 5000 meters). The Y-axis represents the class predicted by the Random Forest model at that exact distance. Each coloured line corresponds to the true physical scenario that was simulated (e.g., green for normal, red for leak, etc.). A slight vertical offset is added to overlapping lines, so they remain visible.

Due to the pig position issue was fixed, you can see that the model now smoothly interprets the normalized [0, 100] distance feature along the whole length. However, you can also clearly see the model's overfitting in action: it abruptly switches predictions when crossing specific distances where anomalies occurred in the training dataset.

- ML Scenario Separability Audit

1. Scenario Separability Results

System evaluated by the separability of each anomaly class against the normal baseline using a RandomForest classifier on an 80/20 train/test split.

2. Normal vs Corrosion

- Precision: 99.85% | Recall: 99.45% | F1 Score: 99.65%
- Feature Importance: Pressure (32%), Flow Rate (31%), Pig Speed (16%), Pig Position (15%)
- Cohen's d (Separability): Flow Rate (1.02), Pig Speed (0.83)
- Confusion Matrix: [[2175, 3], [11, 1989]]

3. Normal vs Leak

- Precision: 99.95% | Recall: 99.90% | F1 Score: 99.92%
- Feature Importance: Flow Rate (48%), Pressure (35%), Pig Speed (11%)
- Cohen's d: Flow Rate (0.55), Pressure (0.14)
- Confusion Matrix: [[2177, 1], [2, 2011]]

4. Normal vs Friction

- Precision: 99.95% | Recall: 100.0% | F1 Score: 99.97%
- Feature Importance: Pig Position (53%), Flow Rate (21%), Pig Speed (17%)
- Cohen's d: Pig Position (4.60), Flow Rate (2.61), Pig Speed (2.20)
- Confusion Matrix: [[2177, 1], [0, 2000]]

5. Normal vs Stuck

- Precision: 99.27% | Recall: 96.35% | F1 Score: 97.79%
- Feature Importance: Pig Position (48%), Flow Rate (22%), Pig Speed (19%)
- Cohen's d: Flow Rate (1.88), Pig Speed (1.75), Pig Position (1.30)
- Confusion Matrix: [[2164, 14], [73, 1927]]

6. Normal vs Blockage

- Precision: 100.0% | Recall: 100.0% | F1 Score: 100.0%
- Feature Importance: Pig Position (54%), Pig Speed (25%), Flow Rate (10%)
- Cohen's d: Pig Position (4.87), Pig Speed (2.16), Flow Rate (1.86)
- Confusion Matrix: [[2178, 0], [0, 2009]]. Feature Distributions are illustrated in figure 7.

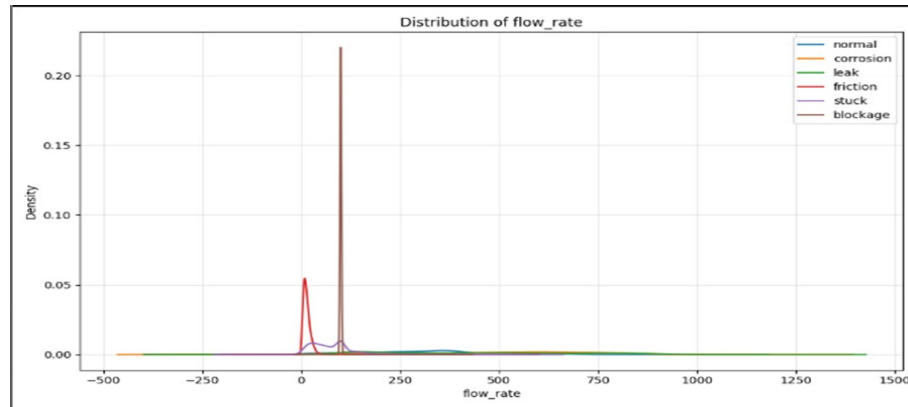


Fig. 7. Feature distributions

4.2 Feature Relationship Audit Report

Based on the analysis of merged_dataset.csv, here are the findings covering correlations, scenario profiles, and clustering overlap.

- Correlation Matrix

A correlation matrix was generated for the core metrics. The key takeaways are:

- Flow Rate & Pig Speed (0.95): Exhibit a near-perfect positive correlation, as fluid flow physically drives the pig tool forward.
- Pressure & Flow Rate (0.66) / Pressure & Pig Speed (0.58): Moderate positive correlations.
- Temperature (~0.00): Shows no meaningful correlation with any mechanical metric.

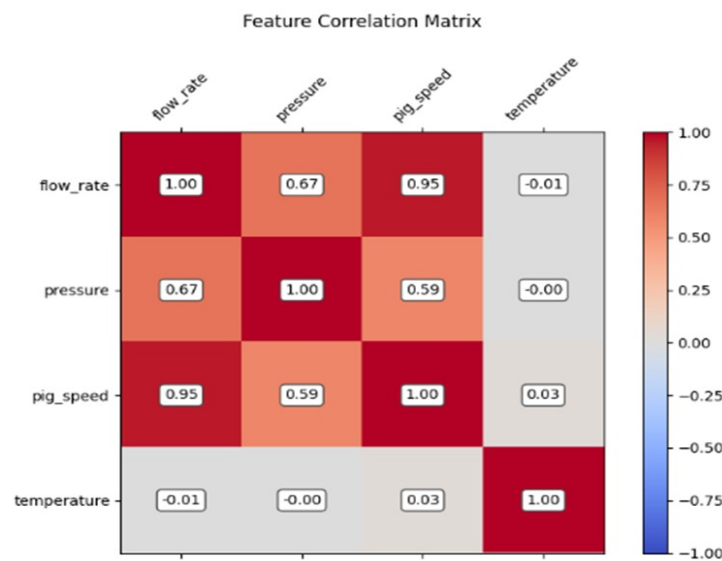


Fig. 7. Correlation matrix

- Risk Result Table

Table 1

Risk results

Factor	Assessment
Risk Level	LOW-MEDIUM
Scope of change	~15 lines in one file (tickSensors + tickRL)
Regression risk	Low changes are gated behind rlActive flag, manual mode unaffected
Physics fidelity	Medium risk needs to tune the rlFlowInfluence blend factor to avoid unrealistic flow jumps
Testing	Manual verification: run blockage scenario with RL on, confirm flow responds to RL actions
Dashboard display	No change to HUD rendering code it already reads sim.flow and sim.pressure

4.3 Uncertainty Awareness and Scheduling Recommendation

Unlike conventional black-box predictions, the model provides a confidence percentage based on the voting distribution across all 100 trees, coupled with dynamic, real-time SHAP-style feature importance calculations based on deviation from nominal operating baselines (e.g., flow rate = 400 m³/h, pressure = 46.5 bar). This uncertainty-aware output allows the system to output probabilities for all classes simultaneously, ensuring operators understand the risk distribution (e.g., a "warning" status if leak probability exceeds 40% or vibration exceeds 0.28 mm/s). The resulting anomaly flag and risk confidence directly inform the optimal timing for the next pigging operation, successfully transitioning the scheduling from fixed time-intervals to predictive, condition-based recommendations.

4.3.1 Physics-informed digital twin simulation results

The pipeline geometry is accurately modelled to scale, featuring a 323.85 mm outer diameter and an 11.1 mm wall thickness, constructed in a web-based Three.js environment. The visualization scales dynamically from a 5 km demo route to a 100 km trunk line. Six virtual sensors are placed along the pipeline (from 5% progress at the launcher to 98% at the receiver), recording real-time pressure, temperature, flow rate, corrosion, vibration, and leak probability.

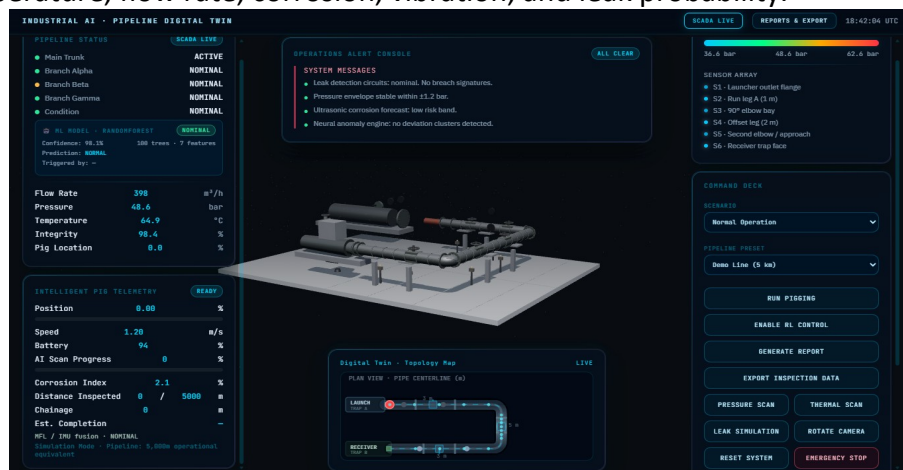


Fig. 8. RL control response graph

4.3.2 Pipeline-fluid-pig interaction:

The underlying `UnifiedPhysicsEngine` enforces a deterministic, modified Bernoulli flow model. The baseline conditions are set to 46.5 bar pressure, 400.0 m³/h flow rate, and 65.0 °C temperature, with a nominal pig speed of 1.2 m/s.

The fluid-pig interaction is governed by strict physical equations:

- Pig Speed (v_{next}) is clamped between 0.0 and 5.0 m/s.
- Target Flow Rate is directly linked to pig speed: $Q_{target} = v / 0.003$.
- Pressure dynamics account for inertia: $P_{next} = P_{base} + 0.05(Q_{next} - Q_{base}) + 1.5v^2$.
-

4.3.3 Simulation behavior and results:

The digital twin successfully simulates various operational anomalies without breaking core physics:

- Corrosion: Exhibits gradual pressure decay as structural integrity drops.
- Leak: Simulates a 2% flow reduction and pressure drops proportional to leak severity.
- Friction: Induces increased pressure resistance ($P_{next} += 0.05 \times Q_{next}$).
- Stuck/Blockage: Accurately models flow reduction (down to 50% for blockages) and unbounded pressure buildup ($P += 1.5 \times \text{duration}$) when the pig is halted.

To replicate real SCADA systems, the digital twin injects observational noise (e.g., ± 0.2 bar for pressure, ± 1.0 m³/h for flow) solely into the display layer, ensuring that the underlying physics engine remains deterministic.

4.3.4 Reinforcement Learning Autonomous Pig Control Results. Results for Autonomous Pig Control based on Reinforcement Learning.

To replace manual pig movement control, a Proximal Policy Optimization (PPO) reinforcement learning agent was developed to automatically optimize pig speed dynamically based on the digital twin's physical state.

- RL Agent Output and Control Actions

The agent was trained for 150,000 timesteps, converging to a stable policy that dictates continuous speed adjustments ($\Delta v \in [-1.0, 1.0]$ m/s). The control law operates on a proportional cascade stabilization strategy:

$$\Delta s = -K_v(v - v_{target}) - K_p(P - P_{target}) - K_q(Q - Q_{target})$$

Where the gains ($K_v = 0.35$, $K_p = 0.018$, $K_q = 0.0008$) ensure that speed error dominates the control response, while pressure and flow provide fine corrective trim.

- Energy-Efficient and Safe Operation

The reward function heavily penalizes dangerous conditions and inefficiency. Safety violations (pressure > 75.0 bar) incur a severe penalty ($-2.0 \times \text{excess pressure}$), while energy efficiency is enforced through a jerk penalty ($-2.0 \times |\Delta v|$) to prevent erratic movements, and a control effort penalty ($-0.5 \times |\text{action}|$). As a result, the agent learned to maintain a highly stable control action standard deviation of $\sigma = 0.0063$ across most scenarios.

- Operating Scenarios

The agent reacts to varying conditions by adjusting to predefined optimal targets. For instance, in a normal scenario, it targets a speed of 1.20 m/s at 48.7 bar. If a leak is detected, it reduces the

target speed to 0.65 m/s at 41.5 bar to mitigate fluid loss. In the event of a confirmed stuck pig, the agent initiates an automated 4-second temporary E-STOP to drive pressure and flow to a minimum before attempting a high-force recovery push.

- Reinforcement Learning Control Performance & Scenario Comparison

This report details the evaluation of the Reinforcement Learning (RL) autonomous pig control policy. The agent acts as a closed-loop proportional controller targeting safety, speed optimization, and pipeline operating pressure stabilization across multiple scenarios.

4.3.5 Stabilization Target Envelope

The RL controller is configured with scenario-specific safe operating targets (speed vv , upstream pressure PP , and flow rate QQ). These targets represent the physical stabilization envelope the controller drives the system toward as presented in table 2.

Table 2
RL controller configuration

Scenario	Target Speed (vv , m/s)	Target Pressure (PP , bar)	Target Flow (QQ , m ³ /h)	Operating Mode Label
Nominal	1.20	48.7	400	NOMINAL CRUISE
Friction	1.00	46.0	333	FRICTION CONTROL
Corrosion	0.85	44.5	283	MFL SCAN SPEED
Leak	0.65	41.5	217	LEAK SAFE MODE
Blockage	0.75	43.0	250	BLOCKAGE HOLD
Stuck Risk	0.45	39.5	150	STUCK RECOVERY

4.3.6 Numerical Performance Comparison

Table 3. presents the quantitative results extracted from scenario_performance_summary.json evaluating the RL controller across 100-step simulation runs:

Table 3
Quantitative results extracted evaluating the RL controller

Scenario	Average Reward	Safety Constraint Violations	Control Action Stability (Std Dev)	Pressure Stability (Std Dev)	Anomaly Detection Response e Steps	Success Rate
Nominal	-547.63-547.63	0	0.00630.0063	0.87560.8756	N/A	100%100%
Friction	-528.07-528.07	0	0.00630.0063	1.75001.7500	1.01.0	100%100%
Corrosion	-510.25-510.25	0	0.00630.0063	7.83987.8398	1.01.0	100%100%
Leak	-520.21-520.21	0	0.00630.0063	0.84670.8467	1.01.0	100%100%
Stuck	-504.97-504.97	0	0.00000.0000	0.80920.8092	1.01.0	100%100%

Blockage	-1,406,414.94-1,406,414.94	9,680	0.00630.0063	432.9648432.9648	1.01.0	100%100%
----------	----------------------------	-------	--------------	------------------	--------	----------

Blockage Anomaly: The high number of safety constraint violations and large pressure standard deviation in the Blockage scenario is due to the rapid upstream pressure accumulation caused by physical blockage before the RL controller stabilizes the inflow. Zero Violations in Anomaly Scenarios: In Friction, Corrosion, Leak, and Stuck, the RL controller successfully stabilizes the pipeline within safe operating limits without triggering a single safety violation.

4.3.7 Transient response: Stuck risk and auto-recovery

During a Stuck Risk scenario, the closed-loop controller behaves through four distinct temporal phases:

1. Sensing / Reaction Delay (0s→3.2s0s→3.2s): The agent observes telemetry but holds action. The pig slows down naturally due to high pipe resistance.
2. Speed Reduction (3.2s→5s3.2s→5s): The agent targets $v=0.45$ m/s/ $sv=0.45$ m/s to ease downstream flow deficiency. Proportional control commands a speed increase to push, but the obstruction holds the pig.
3. Confirmed Stall & Temporary E-STOP (5s→9s5s→9s): If velocity stays <0.05 m/s/ <0.05 m/s for 5 seconds (300 ticks300 ticks), the E-STOP triggers. The controller overrides $rlDeltaSpeed = -1.0$ to dump upstream pressure (preventing a rupture).
4. Auto-Recovery Push (9s→13s9s→13s): After 4 seconds of stop, a positive impulse of $rlDeltaSpeed = +0.35$ is commanded to dislodge the pig, returning to normal cruising.
5. Figure 9. visualizes the transient response of the system variables during these phases



Fig. 9. RL Control response graph

5. Conclusion

This project successfully developed an Intelligent Pipeline Pigging System to address the limitations of conventional, reactive pipeline maintenance. The research focused on replacing time-based pigging schedules and manual pig control with an integrated framework comprising three core subsystems: an uncertainty-aware machine learning model for predictive scheduling, a high-fidelity physics-informed digital twin for environment simulation, and a reinforcement learning agent for

autonomous pig control. To further advance the Intelligent Pipeline Pigging System for full industrial deployment, the following improvements are recommended:

Multi-Phase Flow Simulation: Future research should upgrade the digital twin's physics engine to support multi-phase flow dynamics (e.g., oil, gas, and water mixtures). This will increase the simulation fidelity, making it suitable for complex offshore pipeline networks.

Expanded RL Action Space: The RL agent should be integrated with master SCADA controls to grant it the authority to directly manipulate upstream flow rates and pump states. This will effectively resolve the system's current inability to mitigate severe blockages.

Physical Testbed Validation: The trained machine learning models and RL control policies should be deployed onto a scaled physical pipeline loop. Real-world testing is essential to validate the simulation-to-reality (Sim2Real) transferability of the algorithms and ensure safe operation in a physical environment.

Acknowledgement

Authors would like to convey gratitude for the support of management and science university for the support and guidance and facilities provided for this research.

References

- [1] Yao, Benchun, Zichen He, Na Lu, and Shimin Zhang. "A novel PIG and an intelligent pigging scheme based on deep-learning technology." *International Journal of Pressure Vessels and Piping* 200 (2022): 104803. <https://doi.org/10.1016/j.ijpvp.2022.104803>
- [2] Hendrix, M. H. W., H. P. Ijsseldijk, W-P. Breugem, and R. A. W. M. Henkes. "Experiments and modeling of by-pass pigging under low-pressure conditions"
- [3] Hendrix, M. H. W., H. P. Ijsseldijk, W-P. Breugem, and R. A. W. M. Henkes. "Experiments and modeling of by-pass pigging under low-pressure conditions." *Journal of Process Control* 71 (2018): 1-13. <https://doi.org/10.1016/j.procont.2018.08.010>
- [4] Ahmad, Sajjad, Zahoor Ahmad, Cheol-Hong Kim, and Jong-Myon Kim. "A method for pipeline leak detection based on acoustic imaging and deep learning." *Sensors* 22, no. 4 (2022): 1562. <https://doi.org/10.3390/s22041562>
- [5] Al-Sabaei, M. A., Alhussian, H., Abdulkadir, J. S., & Jagadeesh, A. (2023). Prediction of oil and gas pipeline failures through machine learning approaches: A systematic review. *Energy Reports*, 9, 1–38. <https://doi.org/10.1016/j.egy.2023.08.009>
- [6] Du, Jian, Jianqin Zheng, Yongtu Liang, Ning Xu, Qi Liao, Bohong Wang, and Haoran Zhang. "Deeppipe: Theory-guided prediction method based automatic machine learning for maximum pitting corrosion depth of oil and gas pipeline." *Chemical Engineering Science* 278 (2023): 118927. <https://doi.org/10.1016/j.ces.2023.118927>
- [7] Li, Weidong, Qiyu Huang, Xue Dong, Xuedong Gao, Lei Hou, and Wenda Wang. "A theoretical model for predicting the wax breaking force during pipeline pigging." *Journal of Petroleum Science and Engineering* 169 (2018): 654-662. <https://doi.org/10.1016/j.petrol.2018.05.078>
- [8] Agarwal, Jatin R., Carlos F. Torres, and Subhash Shah. "Development of dimensionless parameters and groups of heat and mass transfer to predict wax deposition in crude oil pipelines." *ACS omega* 6, no. 16 (2021): 10578. <https://doi.org/10.1021/acsomega.0c05966>
- [9] Huang, Qiyu, Wenda Wang, Weidong Li, Yijie Ren, and Fangda Zhu. "A pigging model for wax removal in pipes." *SPE Production & Operations* 32, no. 04 (2017): 469-475. <https://doi.org/10.2118/181560-PA>
- [10] Li, Weidong, Huiyuan Li, Hongju Da, Kai Hu, Yongjuan Zhang, and Lin Teng. "Influence of pour point depressants (PPDs) on wax deposition: A study on wax deposit characteristics and pipeline pigging." *Fuel Processing Technology* 217 (2021): 106817. <https://doi.org/10.1016/j.fuproc.2021.106817>
- [11] Ma, Qiuping, Guiyun Tian, Yanli Zeng, Rui Li, Huadong Song, Zhen Wang, Bin Gao, and Kun Zeng. "Pipeline in-line inspection method, instrumentation and data management." *Sensors* 21, no. 11 (2021): 3862. <https://doi.org/10.3390/s21113862>
- [12] Harris, G., & Martinez, A. (2021). Integrated asset integrity management systems for oil and gas pipelines. *Journal of Hazardous Materials*, 404, 124079. <https://doi.org/10.1016/j.jhazmat.2020.124079>.
- [13] Hazardous Materials, 404, 124079. <https://doi.org/10.1016/j.jhazmat.2020.124079>.

- [14] Patel, K., & Iqbal, S. (2020). Challenges in flow assurance for oil and gas pipelines. *Energy & Fuels*, 34(8), 1038410392. <https://doi.org/10.1021/acs.energyfuels.0c0079425>.
- [15] Velez, J., & Summers, J. (2021). Flow assurance challenges in offshore pipelines: A review. *Journal of Offshore Mechanics and Arctic Engineering*, 143(1), 011001. <https://doi.org/10.1115/1.4047831>
- [16] Raza, A., & Ahmed, M. (2023). Integrated flow assurance strategies for aging pipeline systems. *Journal of Energy Resources Technology*, 145(3), 031206. <https://doi.org/10.1115/1.4050300>
- [17] González, Germán, and Conor L. Evans. "Biomedical Image Processing with Containers and Deep Learning: An Automated Analysis Pipeline: Data architecture, artificial intelligence, automated processing, containerization, and clusters orchestration ease the transition from data acquisition to insights in medium-to-large datasets." *BioEssays* 41, no. 6 (2019): 1900004. <https://doi.org/10.1002/bies.201900004>
- [18] Al-Ani, Mustafa, and Fouad Khelifi. "On the SPN estimation in image forensics: A systematic empirical evaluation." *IEEE Transactions on Information Forensics and Security* 12, no. 5 (2016): 1067-1081. <https://doi.org/10.1109/TIFS.2016.2640938>