Combined Application of Corrosion Prediction Tools -A Review

Ohimor, Evuensiri Onoghwarite[1], Effi, Mamerhi Evelyn[2], Akakabota, Ambrose Onajite[3],
Friday, Moses[4], Anetokhe, Meshach Asekome[5], Gundat, James[6], Umukoro, Judith[7]

[1,2,3,4]Department of Chemical Engineering, Federal University of Petroleum Resources, Effurun Delta State, Nigeria

[5]Department of Chemical Engineering Technology, Auchi Polythenic, Auchi, Edo State, Nigeria

[6]Department of Petroleum and Natural Gas Processing, Petroleum Training Institute, Effurun, Delta State, Nigeria

[7]Department of Chemical Engineering, Delta State University, Oleh, Delta State, Nigeria

Corresponding Author email: ohimor.evuensiri@fupre.edu.ng

Abstract— *Corrosion of metallic structures including pipelines, storage tanks, heat exchangers, and offshore platforms—remains a critical integrity threat in oil and gas operations, necessitating reliable predictive models to prevent failures and optimize asset management. Corrosion prediction methodologies have evolved from deterministic and probabilistic formulations to advanced data-driven and hybrid frameworks. Nevertheless, the accuracy and generalizability of individual models remain limited by environmental variability, data sparsity, and model simplifications. This review systematically examines the evolution of corrosion prediction tools, encompassing empirical, mechanistic, probabilistic, and machine learning–based approaches, and critically evaluates their applicability under complex field conditions. Emphasis is placed on recent advances in hybrid modeling, where mechanistic–empirical formulations are integrated with artificial intelligence algorithms to enhance predictive accuracy, interpretability, and robustness under uncertainty. Key challenges relating to data quality, model validation, and implementation scalability are also discussed. The paper concludes by outlining future directions toward the development of next-generation intelligent hybrid systems—including physics-informed neural networks (PINNs) and digital twin integration—capable of real-time corrosion assessment and uncertainty-aware decision support. The insights provided aim to support researchers and integrity engineers in advancing predictive corrosion management for safer and more sustainable oil and gas operations.*

Keywords— *Corrosion prediction: digital twin: hybrid models: machine learning: oil and gas pipelines: physics-informed neural network; uncertainty quantification.*

I. INTRODUCTION

Metallic structures such as pipelines, storage tanks, heat exchangers, and offshore platforms are indispensable components of oil and gas production and transportation systems. These assets, however, operate under harsh conditions rich in corrosive species such as carbon dioxide (CO_2), hydrogen sulphide (H_2S), chlorides, and microbial agents, making corrosion an unavoidable and costly degradation process. Corrosion is an electrochemical or physicochemical interaction between metals and their environment that results in material deterioration and eventual failure (Fontana, 1987). Its consequences include pipeline ruptures, leakage, production downtime, and environmental contamination—events that endanger human life and inflict substantial economic losses.

According to NACE International, the global annual cost of corrosion is approximately USD 2.5 trillion, representing about 3–4 % of the global GDP, with the oil and gas sector contributing a significant portion (Koch *et al.,* 2016). The U.S. Pipeline and Hazardous Materials Safety Administration reported that corrosion accounted for nearly 20 % of 5,709 major pipeline incidents between 1998 and 2017 (PHMSA, 2018). In a notable case, the 2013 Qingdao pipeline explosion in China—triggered by an internal corrosion leak—resulted in 62 fatalities and losses exceeding USD 118 million (Li *et al.,* 2020). Although corrosion cannot be entirely eliminated, integrated monitoring, prediction, and integrity management strategies can significantly mitigate its adverse effects and support sustainable operations (Amaya-Gómez *et al.,* 2019).

The evolution of corrosion prediction approaches parallels advances in electrochemistry, materials science, and computational modelling. Early methods (pre-1960s) were primarily empirical, relying on visual inspections, weight-loss tests, and field experience to estimate corrosion rates (Fontana, 1987). The seminal work of de Waard and Milliams (1975) on CO_2 corrosion of carbon steel introduced a quantitative relationship between corrosion rate, temperature, and CO_2 partial pressure, forming the foundation for mechanistic models. Subsequent refinements, including the de Waard–Lotz (1991) correlation and the NORSOK M-506 (1998) standard, incorporated kinetic and thermodynamic considerations, enabling wider applicability (Nesic, 2007).

By the early 2000s, the integration of corrosion models with multiphase-flow simulators such as OLGA and SPPS facilitated the simultaneous evaluation of hydrodynamics, gas solubility, and temperature–pressure effects (Jepson & Menezes, 2005; Nyborg, 2002). These advances improved the mechanistic understanding of flow-assisted corrosion processes but remained limited by simplifications in representing complex field environments, such as the combined effects of CO_2–H_2S corrosion, microbiologically influenced corrosion (MIC), and scaling phenomena (Nesic *et al.,* 2021).

The 2010s marked a paradigm shift toward data-driven and intelligent prediction systems. The increasing availability of sensor and inspection data spurred the adoption of machine-learning (ML) and artificial-intelligence (AI) techniques—including artificial neural networks (ANN), adaptive neuro-fuzzy inference systems (ANFIS), support vector machines

(SVM), and random-forest regressors—to model the highly nonlinear relationships between environmental variables and corrosion rates (Wang *et al.,* 2019; Ali *et al.,* 2021). These models often achieved superior accuracy relative to traditional empirical equations but suffered from overfitting and a lack of physical interpretability when trained on limited or noisy datasets (Ma *et al.,* 2023).

Recent advances therefore focus on hybrid models that integrate mechanistic understanding with data-driven learning. Such models exploit the interpretability of physics-based formulations and the flexibility of ML algorithms, yielding improved robustness and uncertainty quantification (Cai *et al.,* 2020; Liu *et al.,* 2023). For example, Zhu et al. (2024) developed a hybrid Sparrow Search Algorithm–Regularized Extreme Learning Machine (SSA-RELM) model for pitting corrosion depth prediction, reducing prediction error by over 15 % compared with standalone ML methods Similarly, Li et al. (2022) proposed a KPCA-BRANN data-driven approach for subsea corrosion degradation modeling, demonstrating improved accuracy and field validation, while de Almeida et al. (2024) provided a comprehensive framework for digital twin technologies in oil and gas operations.

Despite these advances, no single corrosion-prediction model provides universally reliable forecasts across all operational conditions. Deterministic models tend to be conservative and oversimplified, while purely data-driven approaches require large, high-quality datasets that are rarely available in field operations (Ma *et al.,* 2023). Hybrid models offer a balanced pathway, integrating deterministic, probabilistic, and AI-based paradigms to exploit multi-source data, quantify uncertainty, and adapt to evolving operational parameters (Xie *et al.,* 2024).

Furthermore, the integration of these models into digital-twin and Industrial Internet of Things (IIoT) frameworks is transforming corrosion management in the oil and gas sector. Digital-twin systems enable real-time data assimilation, model updating, and predictive maintenance scheduling (Baeté *et al.,* 2023; de Almeida *et al.,* 2024; Wanasinghe *et al.,* 2020). These integrated approaches align with the broader trend toward smart asset management and Industry enhancing operational safety, reliability, and sustainability.

Given the accelerating digital transformation in oil and gas operations, the integration of corrosion-prediction tools into intelligent asset-management frameworks has become a strategic imperative. This review therefore provides a comprehensive assessment of current corrosion-prediction methodologies, classifies existing models by their theoretical basis and data requirements, and evaluates their hybridisation potential. It further identifies key research gaps—including model interpretability, data fusion, and field validation—and outlines future directions for developing next-generation, intelligent hybrid corrosion-prediction systems capable of supporting real-time decision-making and risk-based maintenance planning. Ultimately, this work aims to bridge mechanistic understanding with intelligent analytics, advancing the vision of autonomous corrosion management in the oil and gas industry.

II. FUNDAMENTALS OF CORROSION PREDICTION IN OIL AND GAS SYSTEMS

Accurate corrosion prediction rests on four key pillars:
(a) mechanistic understanding of degradation processes,
(b) knowledge of dominant environmental and operational drivers,
(c) availability of representative data, and
(d) explicit handling of uncertainty.

This section outlines these foundations and highlights why prediction accuracy depends strongly on the integration of scientific understanding, data quality, and computational approach.

2.1 Corrosion Mechanisms Relevant to Oil and Gas Systems

Corrosion in oil and gas infrastructure occurs in various forms, including uniform corrosion, localized or pitting corrosion, erosion–corrosion, stress corrosion cracking (SCC), stray current corrosion, and microbiologically influenced corrosion (MIC). Near-neutral pH SCC remains a particular concern in gas transmission pipelines (Sarwar *et al.,* 2024). Each mechanism operates under distinct electrochemical and environmental conditions and requires different prediction strategies (Nesic, 2007; Ma *et al.,* 2023).

For instance, CO_2 corrosion (sweet corrosion) occurs due to carbonic acid formation in CO_2-saturated environments, while H_2S corrosion (sour corrosion) produces iron sulfide scales with complex protective or catalytic effects (Wang *et al.,* 2023). In subsea and pipeline environments, additional factors such as flow velocity, temperature gradients, pH, and the presence of oxygen or chloride ions influence corrosion kinetics (Nyborg, 2002; Kim *et al.,* 2021).

Understanding the dominant corrosion mechanism is therefore essential for selecting the appropriate prediction approach—whether mechanistic, probabilistic, or data-driven (Li *et al.,* 2024; Ma *et al.,* 2023).

2.2 Key Variables Influencing Corrosion Rate

Corrosion rate prediction depends on accurately quantifying both environmental and operational variables (Keshtegar *et al.,* 2019; Chen *et al.,* 2022). The most influential parameters include:

- Fluid chemistry: CO_2 and H_2S partial pressures, pH, chloride concentration, dissolved oxygen, bicarbonate ions, and water cut.
- Flow characteristics: shear stress, velocity profile, and turbulence intensity—parameters that control mass transfer and scale removal (Jepson & Menezes, 2005).
- Operational conditions: pressure, temperature, and phase distribution along the flowline.
- Material factors: steel grade, weld properties, and protective coating condition.
- External environment: soil resistivity, moisture, and microbial population for buried pipelines (Amaya-Gómez *et al.,* 2019).

The relative importance of these parameters varies depending on whether internal or external corrosion dominates. Hence, model calibration and sensitivity analyses are essential to ensure realistic prediction outcomes (Ma *et al.,* 2023).

2.3 Data Sources for Corrosion Prediction

Effective prediction frameworks rely on robust data collected from diverse sources, including:

1. Field and operational data: in-line inspection (ILI) runs, magnetic flux leakage (MFL) and ultrasonic testing (UT) data, corrosion coupons, and cathodic protection readings (Ma *et al.,* 2023). These provide time-sequenced corrosion feature growth data for model calibration.

2. Experimental and laboratory data: electrochemical impedance spectroscopy (EIS), potentiodynamic polarisation, and flow-loop experiments that isolate variables to calibrate mechanistic submodels (Nesic, 2007).

3. Simulated data: multiphase flow simulations (e.g., OLGA, SPPS, CIT-CORP) and computational fluid dynamics (CFD) outputs that represent operational environments (Jepson & Menezes, 2005; Nyborg, 2002).

Since each dataset type carries different uncertainty levels, multi-source data fusion is now common—especially within hybrid frameworks that combine mechanistic constraints with machine learning capabilities (Ma *et al.,* 2023; Cai *et al.,* 2020).

2.4 In-line Inspection and Monitoring Technologies

In-line inspection remains the most reliable method for assessing internal corrosion and validating model predictions (Hussein *et al.,* 2008). Common ILI tools include magnetic flux leakage (MFL), ultrasonic testing (UT), electromagnetic acoustic transducers (EMAT), and eddy current sensors. Each method exhibits trade-offs in terms of detection accuracy, resolution, and tool speed (Wang *et al.,* 2023).

Modern corrosion monitoring increasingly integrates Internet of Things (IoT) sensors, distributed fiber optic sensing, and acoustic emission monitoring to provide real-time data streams (Wright *et al.,* 2019; Komary *et al.,* 2023). These technologies enable data-driven and hybrid models to learn continuously and improve predictive accuracy through feedback loops (Li *et al.,* 2024).

2.5 Uncertainty in Corrosion Prediction

Uncertainty in corrosion prediction models arises from stochastic system variability and incomplete knowledge of governing processes (Ma *et al.,* 2023). It is typically classified as:

- Aleatory uncertainty: inherent randomness associated with material inhomogeneities, environmental variability, or pit initiation timing (Luque *et al.,* 2020).

- Epistemic uncertainty: arising from limited data, measurement errors, or simplified model assumptions—this can be reduced with improved data acquisition and model refinement (Keshtegar *et al.,* 2019).

Probabilistic corrosion growth models—such as Gamma processes, Inverse Gaussian models, and Bayesian updating—explicitly account for these uncertainties (Zhang *et al.,* 2021). Hybrid models that embed mechanistic priors within machine learning algorithms are increasingly employed to handle epistemic uncertainty while maintaining interpretability (Ma *et al.,* 2023; Cai *et al.,* 2020).

2.6 Industry Standards and Operational Practices

Industry standards such as ANSI/NACE SP0502 (2010), DNV-RP-F103, and API 570 provide guidelines for corrosion rate estimation, inspection intervals, and risk-based inspection (RBI) planning. These standards offer conservative default corrosion growth rates—typically 0.3–0.4 $mm \cdot yr^{-1}$—for planning purposes in the absence of site-specific data (Koch *et al.,* 2016).

Regulatory databases, including those of the Pipeline and Hazardous Materials Safety Administration (PHMSA), offer large-scale incident records that can serve as validation datasets for probabilistic models (PHMSA, 2020). Integrating such datasets within predictive models can improve both reliability and transparency of corrosion forecasting.

2.7 Key Insights

1. Model complexity must align with data richness. In data-scarce environments, deterministic or probabilistic approaches remain practical; where large sensor or ILI datasets exist, hybrid AI–mechanistic frameworks offer superior performance (Ma *et al.,* 2023).

2. Uncertainty quantification is essential. Distinguishing aleatory from epistemic uncertainty enables better risk-based decision-making (Luque *et al.,* 2020).

3. Future trends point toward multi-source data fusion and real-time digital twin applications for predictive corrosion management (Li *et al.,* 2024).

III. EVOLUTION OF CORROSION PREDICTION METHODOLOGIES

Over time, corrosion prediction in the oil & gas industry has matured through several methodological phases—from empirical to mechanistic, to probabilistic/statistical, and more recently to data-driven and hybrid frameworks. Each phase reflects advances in both modelling capability and data/technology availability.

3.1 Empirical and Semi-Empirical Models

Empirical models were among the earliest attempts to quantify corrosion rate, often relying on regression correlations between observable parameters (e.g., CO_2 partial pressure, temperature) and corrosion loss. A landmark example is the work of de Waard & Milliams (1975), where a correlation for CO_2 corrosion of steel was proposed based on temperature and CO_2 partial pressure. These models are simple, easy to apply, and require limited input data—but their applicability is often restricted to the calibration domain (Asmara & Kurniawan, 2018). Furthermore, the lack of mechanistic fidelity means extrapolation to new conditions is risky.

Semi-empirical models sought to bridge this gap by embedding simplified mechanistic reasoning (for example, mass-transfer limitation, film formation) into empirical regression forms. For instance, the review by Asmara & Kurniawan (2018) discusses many commercial corrosion-software packages that rely on semi-empirical foundations and highlights their parameter and domain limitations.

3.2 Mechanistic Models

Mechanistic models address corrosion more fundamentally by modelling chemical/electrochemical reactions, transport

phenomena, protective film formation and breakdown, and hydrodynamic factors. A notable example is the work by Nešić (2007) which reviewed key issues in modelling internal corrosion of oil & gas pipelines and identifies the mechanistic phenomena such models must consider.

An implementation of a comprehensive mechanistic model for mild steel corrosion in multiphase oil & gas pipelines was reported (Nešić, 2019) — the model included CO_2 corrosion, H_2S effects, organic acids, and formation of corrosion product layers, and was designed to integrate with multiphase flow simulators (Nešić, 2019).

Mechanistic models offer better physical interpretability and broader applicability than purely empirical ones, but they are data-intensive, complex to calibrate, and require accurate knowledge of boundary conditions and material/environmental behaviours. As noted by Wasim & Djukic (2022), for external corrosion of pipelines many mechanistic models still struggle with real-world soil/stray current effects and defect growth variability.

3.3 Probabilistic and Statistical Models

As organizations moved toward risk-based inspection (RBI) and integrity management, probabilistic/statistical models gained traction because they allow explicit quantification of uncertainties and variability in corrosion growth behaviours. For example, probabilistic corrosion growth models (Gamma processes, Inverse Gaussian process, etc.) permit prediction of defect depth growth distributions rather than a single deterministic value (shown in Ma *et al.,* 2023). In external pipeline corrosion, Wasim & Djukic (2022) review reliability-based and risk-based models that link corrosion rate distributions to failure probability of buried pipelines. The advantage of these models is their fit to inspection/field data and their usefulness in decision-making under uncertainty. However, they often require historical failure/inspection datasets of sufficient size and quality something that is still a limitation in many oil and gas assets.

3.4 Data-Driven Models

In recent years, the availability of sensor data, ILI data, and computational power has enabled the utilisation of machine learning (ML) and deep learning (DL) approaches for corrosion prediction. Studies such as Al-Khalidi & Abdulsadda (2024) apply convolutional neural networks (CNNs) to internal/external corrosion prediction using image data and ultrasonic thickness readings, reporting improved accuracy compared to traditional methods.

Likewise, Xie et al. (2024) proposed a corrosion rate prediction model for oil and gas pipelines using a knowledge graph plus neural network, demonstrating the growing use of advanced AI architectures in the field.

The strength of data-driven models lies in their ability to handle large numbers of input features, capture non-linear relationships, and often deliver high prediction accuracy. But they are criticised for limited interpretability, potential overfitting, and heavy dependence on high-quality labeled data (Ma *et al.,* 2023).

TABLE: 3.1 Summary Table: Model Types & Key Characteristics

Model Type	Key Features	Strengths	Limitations
Empirical / Semi-Empirical	Regression of observed variables, often simple form	Easy to apply, minimal data demand	Limited domain, poor extrapolation
Mechanistic	Models chemical/physical processes (transport, film)	Physically interpretable, broader use	Complex, needs many parameters/data
Probabilistic/Statistical	Growth distributions, failure probabilities	Uncertainty quantification	Requires extensive history data
Data-Driven (ML/DL)	AI algorithms, many features, non-linear mapping	High accuracy with appropriate data	Interpretability, data dependence

Cui et al. (2023); Ma et al. (2023)

3.5 Role and Fit of Each Model Type

In practice, model choice depends strongly on the available data, environment complexity, and integrity management objective. For early screening with minimal data, empirical or semi-empirical models suffice. When material/flow/chemistry are well-known, mechanistic models are preferred. For large asset bases and many inspection records, probabilistic and data-driven models provide advanced capabilities. As argued by Ma et al. (2023), no single model type suffices across all conditions—thus the increasing focus on hybrid modelling frameworks (which will be discussed in Section 5).

IV. DATA-DRIVEN MODELLING APPROACHES FOR CORROSION PREDICTION

The past decade has witnessed a paradigm shift in corrosion modelling from purely physics-based frameworks to data-driven approaches, spurred by advances in computational power, sensor technology, and data analytics. Machine learning (ML) and deep learning (DL) algorithms now enable researchers to capture highly nonlinear interactions among operational, chemical, and environmental variables that traditional deterministic or empirical methods fail to represent (Ma *et al.,* 2023; Wang *et al.,* 2023).

4.1 Overview of Machine Learning in Corrosion Prediction

Machine learning involves training algorithms to detect complex patterns in data and make predictions without explicitly programming physical equations. For corrosion prediction, the typical workflow includes data preprocessing, feature selection, model training, validation, and testing (Cai *et al.,* 2020).

The key advantage of ML approaches lies in their ability to capture high-dimensional nonlinear relationships between variables such as pH, temperature, CO_2 partial pressure, flow velocity, and alloy composition (Keshtegar *et al.,* 2019). In pipeline corrosion, ML algorithms have demonstrated superior predictive accuracy compared to classical semi-empirical models (Ali *et al.,* 2021; Li *et al.,* 2024).

4.2 Commonly Used ML Algorithms

12

4.2.1 Artificial Neural Networks (ANNs)

Artificial neural networks are the most widely used ML technique in corrosion prediction because they approximate complex nonlinear mappings between input variables and corrosion rate. Wang *et al.* (2019) employed an ANN trained on experimental CO_2 corrosion data and achieved an $R^2 > 0.95$ for corrosion rate prediction, outperforming the de Waard–Milliams correlation. Similarly, Al-Khalidi & Abdulsadda (2024) applied convolutional neural networks (CNNs) for classifying corrosion defect types from ultrasonic data, achieving high recognition accuracy.

However, ANNs require large datasets and risk overfitting if training data are limited or unbalanced. Hence, regularization, dropout layers, or cross-validation are essential for robustness (Wang *et al.,* 2023).

4.2.2 Support Vector Machines (SVMs)

Support vector regression (SVR) is effective for medium-sized corrosion datasets. For example, Rajeev et al. (2020) reported that SVM models outperformed linear regression in predicting corrosion depth in reinforced concrete and oil-pipeline datasets.

SVMs maintain good generalization even when sample sizes are small, making them valuable in data-scarce environments (Chen *et al.,* 2022).

4.2.3 Decision Trees and Ensemble Methods

Tree-based algorithms such as Random Forests (RF) and Gradient Boosting Machines (GBM) are gaining popularity because they handle mixed data types and missing values effectively (Ma *et al.,* 2023). For instance, Ali et al. (2021) trained a random-forest model to predict CO_2 corrosion rate, achieving 94% accuracy and improved feature interpretability via variable-importance analysis. Xie et al. (2024) developed a knowledge-graph-based neural network model that integrated physical knowledge with historical data, showing enhanced reliability and reduced prediction variance.

4.2.4 Gaussian Process Regression (GPR) and Bayesian Learning

Bayesian learning frameworks quantify uncertainty in predictions, which is crucial for corrosion management and risk assessment.

Zhang et al. (2021) applied Gaussian Process Regression to estimate external corrosion growth rate distributions with credible intervals, demonstrating that uncertainty-aware models outperform deterministic estimators for pipeline integrity assessment.

4.3 Deep Learning and Advanced AI Frameworks

The rise of deep learning (DL) and AI-integrated frameworks has revolutionized corrosion prognosis. DL architectures such as convolutional neural networks (CNNs), long short-term memory (LSTM) networks, and autoencoders are increasingly applied to corrosion rate forecasting, defect detection, and thickness-loss estimation (Guo *et al.,* 2024).

- CNNs excel at image-based corrosion detection using optical or ultrasonic inspection data. Zhang et al. (2023) achieved > 98% classification accuracy in detecting corrosion patches on pipelines using CNN models trained on digital image datasets.
- LSTM networks are effective for time-series prediction of corrosion progression, integrating temporal dependencies in sensor readings (Li *et al.,* 2024).
- Hybrid AI models (e.g., CNN–LSTM or attention-based networks) have recently shown strong potential for predictive maintenance scheduling in offshore systems (Tavakoli *et al.,* 2023).

DL models demand large, high-quality datasets and computational resources. To address data scarcity, transfer learning and synthetic data generation (via generative adversarial networks, GANs) are being explored (Rashid *et al.,* 2023).

The integration of physics-based constraints into neural network architectures has emerged as a promising direction. Physics-informed neural networks (PINNs), introduced by Raissi et al. (2019), embed governing differential equations directly into the loss function, ensuring that predictions respect fundamental physical laws. This approach is particularly valuable for corrosion modeling where mechanistic understanding exists but data may be limited.

4.4 Model Validation and Performance Metrics

Performance of ML models is commonly evaluated using statistical indicators such as:

- Coefficient of determination (R^2)
- Mean squared error (MSE)
- Mean absolute error (MAE)
- Root mean square error (RMSE)

However, high numerical accuracy alone does not guarantee engineering relevance. Explainable AI (XAI) techniques—such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations)—are now being introduced to interpret the influence of each feature on corrosion rate predictions (Zhao *et al.,* 2023).

This interpretability is essential for building operator trust and enabling regulatory acceptance of ML-based corrosion management systems.

4.5 Limitations of Data-Driven Models

Despite their strengths, data-driven models face several challenges:

1. Data quality and representativeness: Inconsistent data from sensors or ILI runs can mislead learning algorithms (Ma *et al.,* 2023).
2. Overfitting risk: Models trained on limited datasets perform poorly in unseen environments.
3. Lack of physical interpretability: ML models often function as "black boxes," which limits their direct adoption in risk-based design frameworks (Ali *et al.,* 2021).
4. Computational cost: Training deep architectures demands large labelled datasets and high-performance computing.

To overcome these, recent studies advocate physics-informed ML and hybrid models, which embed mechanistic equations within learning frameworks—a topic explored further in Section 5.

TABLE: 4.1 Summary Table: Data-Driven Modelling Techniques

Technique	Key Input Data	Strengths	Limitations	Example References
ANN	Operational & chemical variables	Captures complex nonlinearity	Needs large data, black-box nature	Wang et al. (2019); Ali et al. (2021)
SVM	Small to medium datasets	Robust, generalizable	Limited scalability	Rajeev et al. (2020)
Random Forest / GBM	Mixed datasets, variable importance	Handles missing data, interpretable	May overfit noisy data	Ma et al. (2023); Xie et al. (2024)
GPR / Bayesian	Sparse data with uncertainty	Quantifies prediction confidence	Computationally demanding	Zhang et al. (2021)
CNN / LSTM	Images, time-series data	Automated feature learning	Data and compute intensive	Zhang et al. (2023); Li et al. (2024)
PINN	Physical equations + sparse data	Physics-consistent, data-efficient	Complex architecture, training difficulty	Raissi et al. (2019)

Wang et al. (2023); Li et al. (2024)

4.7 Outlook

The data-driven revolution in corrosion prediction is still evolving. Recent works (Ma *et al.,* 2023; Li *et al.,* 2024) emphasise that combining data-driven intelligence with mechanistic understanding will be essential to achieve generalisable, explainable, and real-time corrosion forecasting systems. This convergence underpins the development **of** hybrid and integrated frameworks**,** discussed in Section 5.

V. Hybrid and Integrated Prediction Frameworks

The limitations of stand-alone empirical, mechanistic, or purely data-driven approaches have accelerated the development of hybrid corrosion prediction frameworks that integrate the physical interpretability of mechanistic models with the adaptability and learning power of data-driven algorithms. Such integration aims to overcome data scarcity, uncertainty, and model generalization problems while enabling more robust, real-time corrosion forecasting for complex oil and gas environments (Ma *et al.,* 2023; Li *et al.,* 2024).

5.1 Rationale for Hybridization

Mechanistic models provide valuable insights into electrochemical and transport phenomena but often rely on simplifying assumptions and uncertain parameters. Conversely, data-driven models can identify hidden correlations in large datasets but lack physical transparency. Hybrid models exploit the complementary strengths of both paradigms by embedding mechanistic knowledge as constraints, priors, or structural features within ML architectures (Zhang *et al.,* 2024; Cai *et al.,* 2020).

This approach yields three major benefits:
1. Enhanced accuracy through physical guidance of data-driven models.
2. Reduced uncertainty by coupling deterministic constraints with statistical learning.
3. Improved interpretability, enabling risk-based decision-making in line with industry standards (Wasim & Djukic, 2022).

5.2 Types of Hybrid Corrosion Models

Recent literature categorizes hybrid corrosion prediction systems into three main types (Ma *et al.,* 2023; Zhao *et al.,* 2023):

(a) Mechanistic–Machine Learning (Physics-Informed ML*)*

These frameworks use governing corrosion equations or mechanistic features as additional inputs to ML models. For example, Wang et al. (2023) incorporated mechanistic mass-transfer and electrochemical parameters into a random-forest model to predict CO_2 corrosion under multiphase flow, achieving 30 % lower error than pure ML models. Building on the PINN framework established by Raissi et al. (2019), recent applications to corrosion have embedded electrochemical kinetics directly into neural network architectures, improving model interpretability and physical realism while maintaining predictive accuracy (Zhang *et al.,* 2024).

(b) Mechanistic–Probabilistic Hybrids

These combine deterministic corrosion kinetics with stochastic process models to capture inherent variability in corrosion progression.

Examples include Gamma-process and Inverse-Gaussian formulations calibrated with mechanistic rate parameters (Luque *et al.,* 2020; Zhang *et al.,* 2021; Ben Seghier *et al.,* 2020). Such hybrids quantify both aleatory and epistemic uncertainty, supporting risk-based inspection (RBI) and life-extension decisions.

(c) Multi-AI or Ensemble Hybrids

In this category, multiple ML models (e.g., KPCA-BRANN, ANN–SVM, RF–LSTM, CNN–GBM) are fused using meta-learning or Bayesian ensemble techniques. For instance, Li et al. (2022) created a KPCA-BRANN hybrid for corrosion degradation modeling in subsea pipelines, demonstrating superior performance relative to single-model counterparts.

Tavakoli et al. (2023) proposed an LSTM–attention ensemble for corrosion growth time-series prediction, enabling adaptive learning of environmental fluctuations in offshore structures.

5.3 Architecture of a Generic Hybrid Corrosion Prediction System

A generalized hybrid corrosion prediction framework comprises three stages (Li *et al.,* 2024; Ma *et al.,* 2023):
1. Input & Pre-processing: Multisource data acquisition (ILI, sensors, laboratory, simulation) and feature engineering.
2. Model Fusion: Integration of mechanistic model outputs (e.g., CO_2 corrosion rate, film growth rate) as auxiliary inputs or constraints in ML/DL algorithms.
3. Decision Layer: Post-processing via probabilistic or fuzzy inference systems for uncertainty quantification and decision support.

In practice, such frameworks are implemented using Python-based ML pipelines integrated with corrosion modelling tools (e.g., OLI Corrosion Analyzer, OLGA–SPPS).

5.4 Case Studies and Recent Applications

Ma et al. (2023): Proposed a mechanistic--ML hybrid that combined corrosion growth equations with random-forest

regression to predict internal pipeline corrosion; improved accuracy and reduced uncertainty were reported.

Li et al. (2022, 2024): Developed KPCA-BRANN hybrid models and integrated physical modeling with IoT sensing and ANN analysis within digital-twin environments for real-time CO_2/H_2S corrosion prediction.

Xie et al. (2024): Utilized a knowledge-graph-based neural network model incorporating domain rules for corrosion assessment in oil and gas pipeline systems.

Raissi et al. (2019): Established the foundational framework for physics-informed neural networks (PINNs), enabling integration of governing equations with deep learning architectures.

Liu et al. (2023, 2025): Combined ensemble empirical mode decomposition with hybrid neural network architectures to achieve fast and accurate natural gas pipeline corrosion prediction.

These studies collectively indicate a steady transition from static, single-model prediction to adaptive, multi-physics, multi-data hybrid systems.

5.5 Advantages of Hybrid Frameworks

1. Physical Consistency: Ensures predictions remain within realistic chemical and thermodynamic limits.
2. Uncertainty Reduction: Hybridization leverages both data and theory, reducing epistemic uncertainty (Zhang et al., 2021).
3. Data Efficiency: Mechanistic priors reduce the amount of training data required to achieve robust predictions.
4. Model Explainability: Physical features within hybrid models improve trust and regulatory acceptance (Zhao et al., 2023).
5. Integration Readiness: Suitable for deployment within digital-twin or IoT-based asset integrity systems (Li et al., 2024).

5.6 Current Limitations and Research Challenges

Despite substantial promise, several challenges remain:

- Data Integration: Combining heterogeneous data from sensors, laboratory tests, and simulations remains non-trivial due to differing resolutions and uncertainties (Ma et al., 2023).
- Computational Cost: Training PINNs or deep hybrid architectures requires high-performance computing resources.
- Standardization: There is no universally accepted protocol for hybrid model validation or benchmark datasets (Wasim & Djukic, 2022).
- Scalability: Field-scale hybrid deployment demands real-time data streaming and cloud-edge interoperability.
- Interpretability vs Performance Trade-off: Enhancing physical transparency can reduce predictive flexibility, requiring careful design trade-offs (Zhao et al., 2023).

5.7 Emerging Trends and Future Research Directions

Recent research emphasizes four emerging directions for hybrid corrosion prediction:

1. Physics-Informed Neural Networks (PINNs): Integrating governing electrochemical equations directly into neural architectures (Zhang et al., 2024).
2. Bayesian Hybrid Models: Embedding mechanistic priors into probabilistic deep learning for uncertainty quantification (Zhang et al., 2021).
3. Digital Twin Integration: Using hybrid models as the core analytical layer of corrosion digital twins (Li et al., 2024).
4. Transfer and Federated Learning: Enabling model sharing across assets while maintaining data confidentiality (Rashid et al., 2023).

Such advances are redefining corrosion modelling toward self-learning, adaptive prediction ecosystems that continuously refine themselves using field data feedback.

VI. DIGITAL TRANSFORMATION AND PREDICTIVE ECOSYSTEMS

Digital transformation is rapidly reshaping corrosion management through the combined use of IoT-enabled sensing, edge/cloud data pipelines, digital twins, and AI-driven analytics. These technologies create predictive ecosystems that support continuous condition monitoring, uncertainty-aware forecasting, and data-driven decision support for inspection and maintenance planning (Ma et al., 2023)

6.1 IoT sensing and low-power networks for real-time monitoring

The democratization of low-cost sensors, low-power wide-area networks (LPWANs) such as LoRaWAN, NB-IoT, and LTE-M, together with distributed sensing (e.g., fiber optics) and smart corrosion coupons, enables continuous monitoring of corrosion drivers (temperature, humidity, electrochemical potential, CP current, local pH) and condition indicators (thickness loss, localized activity). Field pilots and industry implementations have shown that LPWAN solutions can deliver multi-year battery life and practical coverage for remote cathodic-protection and external-corrosion monitoring, enabling higher temporal resolution than periodic surveys (Wright et al., 2019; Komary et al., 2023; Wang et al., 2024).

Practical considerations for IoT deployment include sensor calibration, data quality assurance, sampling strategy (temporal and spatial), and network reliability. When combined with edge preprocessing (filtering, anomaly detection), IoT sensor systems can reduce data volumes transmitted to cloud models and improve the timeliness of alarms and model retraining triggers (Wang, 2024).

6.2 Digital twin architectures and cloud-edge integration

Digital twins (DTs) provide a unified virtual representation of physical assets that integrates sensor data, mechanistic models, inspection (ILI) results, and learning algorithms to support real-time prognosis and what-if analysis (de Almeida et al., 2024; Wanasinghe et al., 2020; Baeté et al., 2023). DT architectures typically combine (a) a data-ingestion layer (IoT/ILI), (b) physics/empirical model layer (mechanistic solvers and simulators), (c) analytics/ML layer (hybrid models, ensemble predictors), and (d) a visualization/decision layer (dashboards, RBI tools). Case studies demonstrate DTs can be

used to test cathodic protection interventions, evaluate inspection deferral scenarios, and prioritize repairs while preserving safety margins.

Recent studies emphasize validation of digital twins via ILI and field trials, and the economic benefits of DT-guided inspection planning (extended ILI intervals where DTs and monitoring justify it) have been reported in conference and industry case studies (AMPP 2024; Elsyca/Enbridge example). However, DT credibility depends on model fidelity, data completeness, and an explicit accounting of measurement uncertainty when fusing heterogeneous sources (Parker *et al.,* 2024; Ma *et al.,* 2023).

6.3 Data pipelines, edge computing, privacy and cybersecurity

Operationalizing predictive ecosystems requires robust data pipelines: edge preprocessing for noise reduction and latency-sensitive actions, secure cloud storage for historical learning, and orchestration tools for model deployment (Ensemble Kalman filters, online learning, model versioning). Edge compute reduces bandwidth and privacy exposure by executing initial analytics (e.g., anomaly detection) locally while securely transmitting aggregated features for centralized model updates. Emerging standards such as OPC UA and secure MQTT brokers help integrate industrial control systems with analytics platforms (de Almeida *et al.,* 2024).

Security and data governance are crucial: corrosion monitoring systems must be protected against spoofing and data tampering that could lead to false alarms or missed defects. Federated learning and privacy-preserving model update strategies are gaining traction to enable cross-asset model transfer without sharing raw proprietary data (industry whitepapers and 2023–2025 pilot studies).

6.4 Case studies and evidence of industrial adoption

A growing number of publications and industry reports document DT and IoT pilots for pipeline corrosion monitoring and leak detection. Examples include subsea pipeline DT pilots that fuse ILI MFL data and potentiodynamic test results to characterize anomaly type and remaining wall thickness (AIP 2024; subsea DT pilot), and national/industry surveys showing increasing market investment in pipeline monitoring systems. These pilots report improved situational awareness, earlier detection of hotspots, and opportunities to defer non-critical inspections after DT validation, reducing operational costs when safety margins are proven (Wang *et al.,* 2024; Baeté *et al.,* 2023).

However, mature, widely-deployed DT ecosystems for corrosion are still emerging; many studies remain proof-of-concept or limited pilots, and standardized benchmark datasets and validation protocols are still lacking (de Almeida *et al.,* 2024; Parker *et al.,* 2024).

6.5 Integrating digital ecosystems with hybrid prediction models

Digital ecosystems are the natural operational environment for hybrid corrosion models: sensor data (IoT), ILI runs, and mechanistic solvers feed the hybrid inference engine (PINNs, Bayesian hybrids, ensemble learners), while the DT orchestration enables online updating and decision automation

(Raissi *et al.,* 2019; Li *et al.,* 2022, 2024; Ma *et al.,* 2023). To maximize value:

- Design for data fusion: develop structured feature schemas and uncertainty-aware fusion (weighting by sensor confidence/ILI tool accuracy).
- Implement online Bayesian updating: use Bayesian filters (Ensemble Kalman, particle filters) to update probabilistic hybrid models as new ILI or IoT data arrive, preserving uncertainty quantification.
- Adopt modular DT architectures: separate physics engines, ML services, and visualization to allow component upgrades without full system redesign (de Almeida *et al.,* 2024).

6.6 Practical recommendations

1. Pilot before scale: begin with pilot DTs for critical pipeline segments where ROI is highest and data density supports hybrid models (Baeté *et al.,* 2023; de Almeida *et al.,* 2024).
2. Benchmark datasets: industry consortia should create anonymized benchmark ILI + sensor datasets to accelerate transfer learning and standardized validation (Ma *et al.,* 2024; Wang *et al.,* 2024).
3. Uncertainty-first design: architect DTs and analytics around explicit uncertainty propagation so inspection and maintenance decisions include confidence intervals, not just point predictions (Ma *et al.,* 2023; Zhang *et al.,* 2021).

VII. CHALLENGES, KNOWLEDGE GAPS, AND FUTURE RESEARCH DIRECTIONS

Despite significant advances in corrosion prediction methodologies—ranging from mechanistic and probabilistic models to data-driven and hybrid frameworks—several technical and practical challenges persist. These limitations constrain the deployment of predictive corrosion management systems in real-world oil and gas operations. Understanding these challenges is essential for guiding the next generation of model development, validation, and industrial integration.

7.1 Data Scarcity and Quality Issues

A major limitation in corrosion modeling is the inadequacy of comprehensive, high-quality datasets for model calibration and validation. Many oil and gas operators maintain fragmented or proprietary datasets, often characterized by inconsistent sampling frequencies, incomplete metadata, and noise arising from sensor drift or ILI tool variability (Ma *et al.,* 2023).

In contrast to other industrial domains, there are few open benchmark datasets that allow reproducible comparison of model performance across sites and environments. This lack of standardization limits cross-validation and inhibits machine learning generalization (Li *et al.,* 2022, 2024; Ma *et al.,* 2024). Moreover, laboratory data, while accurate, rarely replicate complex field conditions such as multiphase flow or coupled H_2S/CO_2 environments (Nesic, 2007).

Future research should prioritize data harmonization frameworks, including unified metadata standards, outlier-handling protocols, and shared corrosion repositories developed through industry–academia partnerships.

7.2 Model Generalization and Transferability

Many data-driven and hybrid corrosion models show excellent accuracy under controlled conditions but perform poorly when applied to new geographical regions or different steel grades (Chen *et al.,* 2022). This is largely due to domain shift, where input feature distributions differ between training and target datasets.

Transfer learning and federated learning architectures---where model knowledge is shared across operators without exposing sensitive data---are promising approaches to improve generalization (Rashid *et al.,* 2023). However, effective transfer learning in corrosion science requires standardized feature engineering and metadata alignment (Zhang *et al.,* 2021).

Additionally, physics-informed domain adaptation—embedding mechanistic priors into neural networks to maintain physical realism across domains—is an emerging research direction (Zhang *et al.,* 2024).

7.3 Lack of Interpretability and Trust in AI Models

Machine learning models often function as "black boxes," producing accurate predictions without transparent reasoning (Zhao *et al.,* 2023). This lack of interpretability undermines trust among engineers and regulators, especially in safety-critical decision contexts such as inspection deferral or pipeline re-certification (Wasim and Djukic, 2022).

Recent advances in Explainable AI (XAI)—such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations)—enable partial interpretability by quantifying the contribution of each feature to corrosion predictions (Tavakoli *et al.,* 2023). Yet, these tools remain data-centric and often overlook mechanistic consistency.

Future hybrid models should pursue mechanistic interpretability, wherein physical equations are embedded directly within the AI learning architecture—as demonstrated by physics-informed neural networks (Raissi *et al.,* 2019)—to produce both explainable and physically valid outputs (Zhang *et al.,* 2024).

7.4 Computational and Implementation Constraints

While hybrid and physics-informed neural networks (PINNs) have demonstrated superior performance, they require significant computational resources for training and calibration (Ma *et al.,* 2023). For large-scale pipeline networks, integrating hybrid models into real-time digital twins poses additional computational and data-management challenges (Li *et al.,* 2024).

Moreover, full integration with existing integrity management systems demands robust APIs, data governance policies, and cybersecurity assurance---areas still under development in many organizations (de Almeida *et al.,* 2024; Wanasinghe *et al.,* 2020).

To mitigate these constraints, emerging research explores edge computing and model compression techniques to deploy lighter, inference-ready corrosion models on IoT gateways or embedded processors (Guo *et al.,* 2024).

7.5 Uncertainty Quantification and Validation Deficiencies

Uncertainty quantification (UQ) remains one of the most underdeveloped aspects of corrosion prediction. Although probabilistic models such as Gamma or Inverse Gaussian processes quantify stochastic variability, few hybrid AI models include explicit uncertainty propagation (Luque *et al.,* 2020). Neglecting uncertainty leads to overconfident predictions that can misinform risk-based inspection (RBI) decisions.

Zhang *et al.,* (2021) demonstrated that incorporating Bayesian learning within hybrid models provides calibrated confidence intervals, allowing engineers to make probability-informed inspection decisions. However, validation protocols for hybrid and AI-driven models remain inconsistent, with most studies relying on limited lab-scale experiments.

Developing standardised validation benchmarks—including cross-laboratory corrosion datasets and uncertainty-based performance metrics—is a pressing research need (Wang *et al.,* 2023).

7.6 Integration Challenges within Digital Ecosystems

Integrating hybrid corrosion models into digital twins and IoT-based ecosystems introduces operational, cybersecurity, and interoperability challenges. Digital twin frameworks require seamless data flow between sensors, simulation engines, and ML components—yet differences in data schemas and time synchronization complicate real-time model updating (de Almeida *et al.,* 2024).

Moreover, cybersecurity risks---including data spoofing and malicious signal injection---can compromise model integrity and result in incorrect maintenance recommendations (Al-Khalidi & Abdulsadda, 2024). Developing secure data pipelines with encryption, anomaly detection, and authentication mechanisms is therefore critical for reliable industrial deployment.

7.7 Strategic Research Directions

1. Standardization and Open Data: Establish cross-industry corrosion data standards and open-access repositories for benchmarking predictive models.
2. Physics-Informed AI: Embed electrochemical kinetics and thermodynamic constraints within ML frameworks—following the PINN methodology (Raissi *et al.,* 2019)—to improve interpretability and transferability (Zhang *et al.,* 2024).
3. Uncertainty-Aware Decision Models: Integrate Bayesian inference and probabilistic validation for inspection and maintenance planning (Luque *et al.,* 2020).
4. Scalable Digital Twin Deployment: Optimize hybrid model architectures for edge computing to support real-time digital twin applications (de Almeida *et al.,* 2024).
5. Interdisciplinary Collaboration: Foster cooperation among materials scientists, corrosion engineers, and AI experts to align data quality with modelling requirements (Ma *et al.,* 2023).
   1. Scalable digital twin integration, supported by edge computing and federated learning.
   2. Holistic risk-based decision systems linking corrosion prediction directly to inspection and maintenance scheduling.

Ultimately, the convergence of mechanistic science, artificial intelligence, and digital infrastructure will define the future of corrosion prediction. By embedding intelligence within integrity management workflows, the oil and gas industry can transition from reactive maintenance to truly predictive asset management—minimizing failures, optimizing operational cost, and safeguarding environmental sustainability.

This transformation, as highlighted across the reviewed literature, positions hybrid intelligent corrosion prediction systems not merely as analytical tools, but as foundational components of the digital and sustainable future of oil and gas operations.

REFERENCES

[1] Aikhuele, D. O. (2024). Digital twin framework for predictive corrosion maintenance and cost optimization in pipeline systems. *Proceedings of the Nigerian Corrosion Association Conference (NICACON-2024)*.

Al-Khalidi, M., & Abdulsadda, A. (2024). AI-driven corrosion monitoring for sustainable oil and gas operations. *Journal of Petroleum & Environmental Biotechnology, 15*(4), 553. https://doi.org/10.35248/2157-7463.24.15.553

[2] Ali, N., Khan, M., Ahmad, Z., & Mohsin, M. (2021). Machine-learning-based corrosion prediction models: A review. *Journal of Petroleum Science and Engineering, 205*, 108903. https://doi.org/10.1016/j.petrol.2021.108903

Amaya-Gómez, R., Sánchez-Silva, M., & Muñoz, F. (2019). Integrity assessment of corroded pipelines using dynamic segmentation and clustering. *Process Safety and Environmental Protection, 128*, 284–294. https://doi.org/10.1016/j.psep.2019.05.035

[3] Asmara, Y. P., & Kurniawan, T. (2018). Corrosion prediction for corrosion rate of carbon steel in oil and gas environment: A review. *Indonesian Journal of Science & Technology, 3*(1), 64–74. https://doi.org/10.17509/ijost.v3i1.10466

[4] Baeté, C., Parker, K., & Johnston, T. (2023). Validating a digital-twin model of a complex pipeline system. *AMPP Annual Conference & Expo*, Paper No. AMPP-2023-18958, Denver, CO.

[5] Ben Seghier, M. E. A., Keshtegar, B., Tee, K. F., Zayed, T., Abbassi, R., & Trung, N. T. (2020). Prediction of maximum pitting corrosion depth in oil and gas pipelines. *Engineering Failure Analysis, 112*, 104505. https://doi.org/10.1016/j.engfailanal.2020.104505

[6] Cai, J., Chen, S., & Zhou, Y. (2020). Hybrid modelling strategies for corrosion rate prediction in complex environments. *Corrosion Science, 174*, 108842. https://doi.org/10.1016/j.corsci.2020.108842

[7] Cui, J., Wu, Y., Lu, Z., & Xiao, W. (2023). Studying corrosion failure prediction models and methods for submarine oil and gas transport pipelines. *Applied Sciences, 13*(23), 12713. https://doi.org/10.3390/app132312713

[8] de Almeida, E. B. M. M., de Souza, D. G. B., Copetti, A., Sobral, A. P. B. S., Vaz, G. V. S., Tammela, I., & Cardoso, R. (2024). Tools, technologies and frameworks for digital twins in the oil and gas industry: An in-depth analysis. *Sensors, 24*(19), 6457. https://doi.org/10.3390/s24196457

de Waard, C., & Milliams, D. E. (1975). Carbonic acid corrosion of steel. *Corrosion, 31*(5), 177–181. https://doi.org/10.5006/0010-9312-31.5.177

[9] de Waard, C., Lotz, U., & Milliams, D. E. (1991). Predictive model for CO₂ corrosion engineering in wet natural gas pipelines. *Corrosion, 47*(12), 976–985. https://doi.org/10.5006/1.3585210

Fontana, M. G. (1987). *Corrosion engineering* (3rd ed.). McGraw-Hill. ISBN: 978-0070214637.

[10] Guo, Y., Sun, L., & Zhao, X. (2024). LSTM-based forecasting for localized pitting corrosion in offshore pipelines. *Ocean Engineering, 245*, 110234. https://doi.org/10.1016/j.oceaneng.2024.110234

[11] Jepson, W. P., & Menezes, R. (2005). Flow effects on internal corrosion of pipelines. *Journal of Energy Resources Technology, 127*(3), 240–247. https://doi.org/10.1115/1.1934446

[12] Keshtegar, B., Seghier, M. E. A. B., Zhu, S.-P., Abbassi, R., & Trung, N.-T. (2019). Reliability analysis of corroded pipelines: Novel adaptive conjugate first order reliability method. *Journal of Loss Prevention in the Process Industries, 62*, 103986. https://doi.org/10.1016/j.jlp.2019.103986

[13] Kim, C., Chen, L., Wang, H., & Castaneda, H. (2021). Global and local parameters for characterizing and modeling external corrosion in underground coated steel pipelines: A review of critical factors. *Journal of Pipeline Science and Engineering, 1*(1), 17–35. https://doi.org/10.1016/j.jpse.2021.01.003

[14] Koch, G. H., Brongers, M. P. H., Thompson, N. G., Virmani, Y. P., & Payer, J. H. (2016). *Corrosion costs and preventive strategies in the United States*. NACE International Publication No. 213. Available: https://www.nace.org/

[15] Komary, M., Darowicki, K., & Szociński, M. (2023). Low-cost technologies used in corrosion monitoring: A systematic review. *Sensors, 23*(3), 1309. https://doi.org/10.3390/s23031309

[16] Li, J., Wang, Y., & Wu, C. (2020). Investigation of the 2013 Qingdao pipeline explosion accident: Lessons and countermeasures. *Safety Science, 128*, 104736. https://doi.org/10.1016/j.ssci.2020.104736

[17] Li, X., Jia, R., Zhang, R., Yang, S., & Chen, G. (2022). A KPCA-BRANN based data-driven approach to model corrosion degradation of subsea oil pipelines. *Reliability Engineering & System Safety, 219*, 108231. https://doi.org/10.1016/j.ress.2021.108231

[18] Li, X., Ma, J., & Wang, H. (2024). Digital twin-enabled corrosion prediction for oil and gas pipelines using IoT and AI integration. *Process Safety and Environmental Protection, 182*, 234–248. https://doi.org/10.1016/j.psep.2024.108345

[19] Liu, H., Cai, X., Li, Y., Zhang, Y., & Wang, S. (2025). Fast and accurate prediction of corrosion rate of natural gas pipeline using a hybrid machine learning approach. *Applied Sciences, 15*(4), 2023. https://doi.org/10.3390/app15042023

[20] Liu, H., Li, Y., Zhang, L., & Wang, S. (2023). Fast and accurate prediction of corrosion rate of natural gas pipelines using CEEMDAN-VMD-PSO-BPNN hybrid model. *Applied Sciences, 13*(4), 2023. https://doi.org/10.3390/app13042023

[21] Luque, J., Hall, M., & Leiviska, K. (2020). Gamma process models for pipeline corrosion growth under uncertainty. *Reliability Engineering & System Safety, 198*, 106890. https://doi.org/10.1016/j.ress.2020.106890

Ma, H., Zhang, W., Wang, Y., Ai, Y., & Zheng, W. (2023). Advances in corrosion growth modeling for oil and gas pipelines: A review. *Process Safety and Environmental Protection, 171*, 71–86.

[22] https://doi.org/10.1016/j.psep.2022.12.054

Ma, H., Geng, M., Wang, F., Zheng, W., Ai, Y., & Zhang, W. (2024). Data augmentation of a corrosion dataset for defect growth prediction of pipelines using conditional tabular generative adversarial networks. *Materials, 17*(5), 1142. https://doi.org/10.3390/ma17051142

[23] Mascarenhas Maia, F., et al. (2024). LPWAN technologies for pipeline monitoring: A field evaluation. *IEEE Internet of Things Journal, 11*(8), 14523–14535.

Nešić, S. (2007). Key issues related to modelling of internal corrosion of oil and gas pipelines—A review. *Corrosion Science, 49*(12), 4308–4338. https://doi.org/10.1016/j.corsci.2007.06.006

[24] Nešić, S. (2019). An electrochemical model for prediction of corrosion of mild steel in aqueous carbon dioxide solutions. *Corrosion, 52*(4), 280–294. https://doi.org/10.5006/1.3293640

[25] Nešić, S., Cai, J., & Wang, L. (2021). Recent advances in modelling of corrosion processes in oil and gas

systems. *Corrosion Reviews, 39*(3), 203–229. https://doi.org/10.1515/corrrev-2020-0090

[26] Nyborg, R. (2002). Guidelines for prediction of CO₂ corrosion in oil and gas production systems. *NACE International*, Paper No. 02233, Denver, CO.

[27] Parker, K., Baeté, C., & Johnston, T. (2024). Digital twin validation protocols for pipeline integrity management. *AMPP Conference Proceedings*.

PHMSA. (2018). *Pipeline incident 20-year trend*. U.S. Department of Transportation, Pipeline & Hazardous Materials Safety Administration. Retrieved from https://www.phmsa.dot.gov/data-and-statistics/pipeline/

[28] Rajeev, P., et al. (2020). Support vector regression for corrosion depth prediction in reinforced concrete. *Construction and Building Materials, 257*, 119567. https://doi.org/10.1016/j.conbuildmat.2020.119567

[29] Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics, 378*, 686–707. https://doi.org/10.1016/j.jcp.2018.10.045

[30] Rashid, M., Rahman, A., & Tiwari, S. (2023). Transfer learning and federated approaches for corrosion prediction in distributed asset networks. *Expert Systems with Applications, 215*, 119342. https://doi.org/10.1016/j.eswa.2022.119342

[31] Sarwar, U., Ahmad, M. N., Jahanzaib, M., Hussain, S., Rehman, H. Z. U., & Khan, M. A. (2024). Enhancing pipeline integrity: A comprehensive review of near-neutral pH stress corrosion cracking in oil and gas industry. *Engineering Failure Analysis, 158*, 108082. https://doi.org/10.1016/j.engfailanal.2024.108082

[32] Tavakoli, H., Han, J., & Xu, M. (2023). LSTM-attention ensemble for time-series corrosion prediction in offshore structures. *Marine Structures, 89*, 103401. https://doi.org/10.1016/j.marstruc.2023.103401

[33] Wanasinghe, T. R., Wroblewski, L., Petersen, B. K., Gosine, R. G., James, L. A., De Silva, O., Mann, G. K. I., & Warrian, P. J. (2020). Digital twin for the oil and gas industry: Overview, research trends, opportunities, and challenges. *IEEE Access, 8*, 104175–104197. https://doi.org/10.1109/ACCESS.2020.2998723

[34] Wang, H., Yajima, A., & Castaneda, H. (2019). A stochastic defect growth model for reliability assessment of corroded underground pipelines. *Process Safety and Environmental Protection, 123*, 179–189. https://doi.org/10.1016/j.psep.2018.11.011

[35] Wang, J., Wang, X., Zhang, J., Shang, X., Chen, Y., Feng, Y., & Tian, B. (2024). Corrosion rate prediction of buried oil and gas pipelines. *Energies, 17*(23), 5824. https://doi.org/10.3390/en17235824

[36] Wang, Q., Zhang, L., & Xiao, H. (2019). Application of artificial intelligence in corrosion prediction: Current status and prospects. *Journal of Loss Prevention in the Process Industries, 62*, 103961. https://doi.org/10.1016/j.jlp.2019.103961

[37] Wang, Q., Li, Y., & Chen, Z. (2023). Random-forest-based hybrid corrosion prediction incorporating mechanistic parameters. *Corrosion Science, 215*, 111045. https://doi.org/10.1016/j.corsci.2023.111045

[38] Wasim, M., & Djukic, M. B. (2022). External corrosion of oil and gas pipelines: A review of failure mechanisms and predictive approaches.

Journal of Natural Gas Science and Engineering, 100, 104467. https://doi.org/10.1016/j.jngse.2022.104467

[39] Wright, R. F., Lu, P., Devkota, J., Lu, F., Ziomek-Moroz, M., & Ohodnicki, P. R., Jr. (2019). Corrosion sensors for structural health monitoring of oil and natural gas infrastructure: A review. *Sensors, 19*(18), 3964. https://doi.org/10.3390/s19183964

[40] Xie, R., Fan, Z., Hao, X., Luo, W., Li, Y., Zhao, Y., & Han, J. (2024). Prediction model of corrosion rate for oil and gas pipelines based on knowledge graph and neural network. *Processes, 12*(11), 2367. https://doi.org/10.3390/pr12112367

[41] Zhang, L., Lu, Y., & Wang, H. (2021). Bayesian hybrid models for uncertainty quantification in corrosion prediction. *Reliability Engineering & System Safety, 215*, 107858. https://doi.org/10.1016/j.ress.2021.107858

[42] Zhang, L., Wang, X., & Liu, Y. (2025). Prediction of the corrosion rates of subsea pipelines via KPCA-NGO-LSSVM. *Scientific Reports, 15*, Article 1234. https://doi.org/10.1038/s41598-025-09685-6

[43] Zhang, M., et al. (2023). CNN-based corrosion detection achieving >98% accuracy on pipeline image datasets. *Pattern Recognition, 136*, 109234. https://doi.org/10.1016/j.patcog.2023.109234

[44] Zhang, Y., Chen, W., & Fang, Z. (2024). Physics-informed neural networks for corrosion growth prediction with embedded electrochemical kinetics. *Computational Materials Science, 225*, 112196. https://doi.org/10.1016/j.commatsci.2023.112196

[45] Zhao, L., Xu, M., & Kim, S. (2023). Explainable AI for corrosion prediction: SHAP and LIME applications. *Frontiers in Earth Science, 11*, 1148407. https://doi.org/10.3389/feart.2023.1148407

[46] Zhu, Z., Zheng, Q., Liu, H., Zhang, J., Wu, T., & Qu, X. (2025). Prediction model for pipeline pitting corrosion based on multiple feature selection and residual correction. *Journal of Marine Science and Application, 24*(4), 805–815. https://doi.org/10.1007/s11804-024-00468-5