Scientific Research Publishing

**Open Journal of Applied Sciences, 2026, 16(1), 383-400**
https://www.scirp.org/journal/ojapps
ISSN Online: 2165-3925
ISSN Print: 2165-3917

# AI Assisted Material Selection Framework for Corrosion Resistant Steels in Onshore Oil and Gas Pipelines

**Stephen Oluwatosin Okegbenro[1], Opeyemi Bosede Daniyan[2,3*], Kunle Michael Oluwasegun[4], Olakanmi Adekunle Adewara[3], Bodunde Odunola Akinyeyemi[2], Ganiyu Adesola Aderounmu[2]**

[1]Department of Materials Science and Engineering, Obafemi Awolowo University, Ile-Ife, Nigeria
[2]Department of Computer Science and Cybersecurity, Obafemi Awolowo University, Ile-Ife, Nigeria
[3]Information Technology and Communication Unit, Obafemi Awolowo University, Ile-Ife, Nigeria
[4]Department of Mechanical and Manufacturing Engineering, University of Manitoba, Winnipeg, Canada
Email: *odaniyan@oauife.edu.ng

**How to cite this paper:** Okegbenro, S.O., Daniyan, O.B., Oluwasegun, K.M., Adewara, O.A., Akinyeyemi, B.O. and Aderounmu, G.A. (2026) AI Assisted Material Selection Framework for Corrosion Resistant Steels in Onshore Oil and Gas Pipelines. *Open Journal of Applied Sciences*, **16**, 383-400.
https://doi.org/10.4236/ojapps.2026.161024

**Received:** December 2, 2025
**Accepted:** January 24, 2026
**Published:** January 27, 2026

## Abstract

Corrosion is one of the most challenging problems that affects the safety and durability of onshore pipelines. Corrosion-resistant steels play a pivotal role in ensuring long lasting function, cost reduction in upkeep, and prevention of major breakdowns. The traditional materials selection methods often depend on hands-on testing, expert experience and the use of a trial-and-error process. The rapid advancement of machine learning and artificial intelligence (AI) driven techniques is revolutionizing materials discovery, property prediction, and material design by minimizing human intervention and accelerating scientific progress. In the advent of the digital revolution, Artificial Intelligence (AI) has emerged as a pivotal tool in various domains, including materials design and discovery. This paper aims to develop and facilitate how AI is used to interpret Ashby's chart and merge databases of materials together to select materials for corrosion resistant steels in onshore oil and gas pipelines. This paper proposes an AI Assisted material selection framework specifically designed to identify optimal corrosion resistant steels for onshore pipelines oil and gas pipelines by improving forecast accuracy, reducing uncertainty, enabling informed material selection, and strengthening longterm pipeline integrity in corrosive operating environments. By leveraging machine learning (ML) techniques such as Gradient Boosting Machines (GBM), Principal Component Analysis (PCA) for dimensionality reduction, and Neural Networks. The framework intelligently analyzes vast datasets encompassing operational parameters, environmental conditions, and historical corrosion rates.

**Keywords**

Corrosion, Pipelines, Machine Learning, Artificial Intelligence

## 1. Introduction

Selecting the optimal material in manufacturing for onshore pipelines is becoming more and more complex and time consuming due to the large number of options which currently exist. Many decision techniques have been developed but most of them are only intended to be used in certain fields or contexts and, in general, there is not a standardized way to proceed [1] [2]. The transportation of crude oil and natural gas through pipelines remains one of the most effective and economical means of energy delivery worldwide. However, the long-term integrity of these systems is constantly threatened by corrosion, which has been identified as one of the most critical causes of pipeline failures [3]. Corrosion reduces wall thickness, promotes stress corrosion cracking, and ultimately compromises burst pressure capacity, leading to leaks, ruptures, and catastrophic failures. For onshore oil and gas pipelines in particular, where exposure to harsh environments and operational stresses is common, corrosion continues to pose a significant safety, environmental, and economic risk [4].

## 2. Methodology

### 2.1. Corrosion Challenges in Onshore Pipelines

Onshore oil and gas pipelines face significant corrosion challenges due to environmental factors like soil conditions, moisture, and corrosive fluids. Common corrosion types include $CO_2$ corrosion, causing uniform thinning in sweet environments; $H_2S$ corrosion, leading to sulfide stress cracking in sour service; microbiologically influenced corrosion (MIC), driven by bacterial activity in low flow areas; pitting corrosion, resulting in localized material loss; and stress corrosion cracking (SCC), worsened by tensile stresses and corrosive media [5]-[8]. These mechanisms threaten pipeline integrity, risking leaks or failures [9]. Material choices typically include carbon steels for cost effectiveness in mild conditions, stainless steels for improved resistance to pitting and general corrosion, and duplex steels for enhanced strength and corrosion resistance in aggressive environments with high chloride or $H_2S$ content [5] [6]. Brief description on basic of corrosion are presented in **Table 1**.

### 2.2. Traditional Material Selection Approaches

Traditional material selection approaches typically involve a three-step methodology: screening materials using constraints, ranking them using performance indices, and finally seeking supporting information for top candidates. These methods rely on structured data and include strategies such as quantitative analysis, questionnaire-based expertise capture, and case-based analogy [11]. Traditional

**Table 1.** Brief on basic types of corrosion [10].

| | | |
|---|---|---|
| 1 | Sour Corrosion | is the degradation of a metal when it is in contact with aqueous $H_2S$. |
| 2 | Sweet Corrosion | is the metal degradation in aqueous $CO_2$. |
| 3 | $H_2S$ and $CO_2$ acid fumes | the role of the Oxygen in corrosion is not only accelerating the metal anodic oxidation, but also heightens the $H_2S$ and $CO_2$ acid fumes. |
| 4 | Crevice Corrosion | is a localized corrosion occurring in crevices where fluid becomes stagnant. |
| 5 | Galvanic Corrosion | occurs when one metal meets another conducting metal with different electrochemical potentials in a corrosive medium. |
| 6 | Erosion Corrosion | is the material deterioration brought on by surface oxidation along with mechanical wear due to the impact of solid particles,<br>liquid, or a combination of both processes. |
| 7 | Stress Corrosion Cracking (SCC) | is a delayed, environmentally driven crack propagation interplay of mechanical stress and corrosion reactions. |
| 8 | Microbiologically Influenced Corrosion (MIC) | can be defined as a type of corrosion where the corrosion of materials is brought on and/or accelerated by the actions of microorganisms:<br>i. MIC occurs via several mechanisms and types of microbes.<br>ii. MIC is an electrochemical process among electron donor, electron acceptor, and water. Carbon supply, energy source, and bacteria must exist.<br>iii. MIC may affect the corrosion process of typical grades of stainless steel.<br>iv. MIC frequently occurs on welds and heat affected zones in stagnant or slowly moving waters. |
| 9 | Chloride induced localized corrosion | Lowerballoyed stainless steels, e.g., 304, are insufficiently resistant to corrosion for prolonged exposure to seawater and high chloride concentration environments, and are vulnerable to chloride induced localized corrosion. |

material selection for corrosion resistant steels in pipelines relies on performance testing, such as immersion or electrochemical tests to measure corrosion rates, cost benefit analysis to balance initial and lifecycle costs, and empirical models based on standards like NACE MR0175 for sour service [12]. These methods evaluate alloy composition, mechanical properties, and environmental compatibility to select materials like carbon or stainless steels [13]. However, limitations include time intensive testing, data gaps from incomplete field histories, and subjectivity in interpreting results or tradeoffs, often leading to conservative designs or overlooked risks [14].

### Performance Testing, Cost Benefit Analysis, Empirical Models

[16] presents an empirical evaluation of test case prioritization techniques within Junit based Java environments. The study demonstrates that prioritization significantly enhances fault detection rates, particularly for techniques incorporating feedback. Performance was measured via the APFD metric, with method level and block level techniques showing notable improvements over random or untreated test orders. A cost benefit analysis reveals that prioritization is most effective in incremental or continuous testing contexts, where frequent test executions amplify time savings. The authors introduce cost models differentiating between batch and incremental testing processes, highlighting that savings. **Figure 1** shows

the organization of structured data of the sort generally found in handbooks and data sheets. While the different stages for materials and process selection in design and the conceptual tools to carry them out are presented in **Figure 2**.

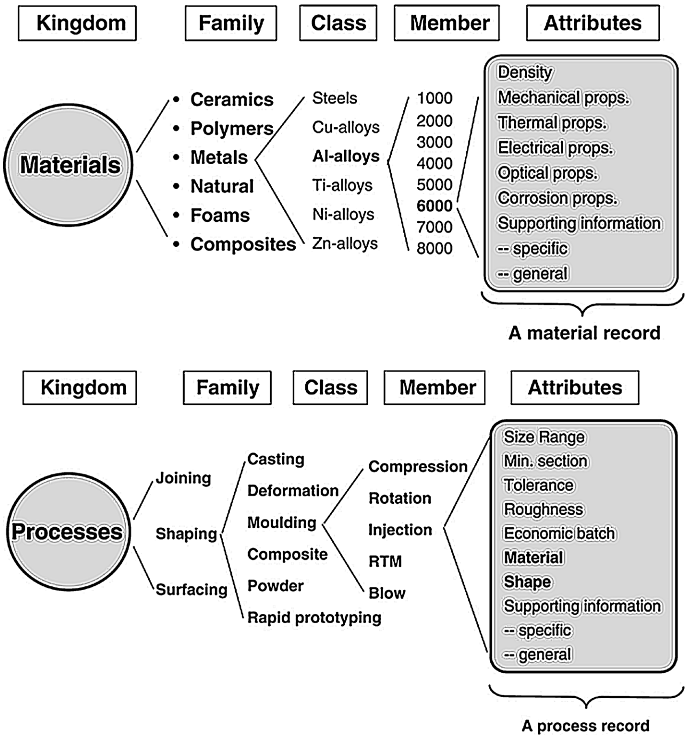

Figure 1. The organization of structured data of the sort generally found in handbooks and data sheets. The material or process is indexed by name; the information listed under the name gives numeric values for material properties, a ranking of performance under standard conditions and yes or no categorization of material–process compatibility [11].

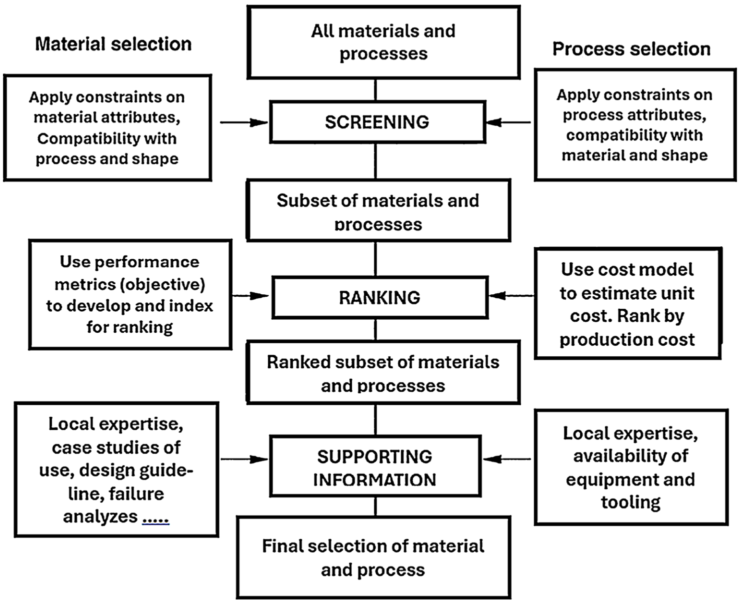

Figure 2. The different stages for materials and process selection in design and the conceptual tools to carry them out [15].

depend on organizational cost structures and test suite characteristics. The empirical models used grounded findings in real world Java programs and seeded faults, offering practical insights into tradeoffs between prioritization overhead and early fault detection benefits. The study underscores the context dependent value of prioritization and provides a framework for assessing its applicability in industrial settings. a corrosion management case study comparing the lifecycle costs of carbon steel and corrosion-resistant alloy (CRA) pipelines in aggressive environments is also utilized. Although CRAs have higher initial material and installation costs, their superior resistance to corrosion significantly reduces maintenance, inspection, downtime, and failure-related expenses over the service life of the asset, resulting in a lower total lifecycle cost.

## 2.3. AI Assisted Material Selection

AI assisted material selection represents a transformative approach to identifying corrosion resistant steels for onshore oil and gas pipelines, leveraging advanced computational techniques to overcome the limitations of traditional methods. By integrating machine learning (ML) and deep learning (DL) algorithms, this approach enables rapid, data driven predictions of material performance under complex corrosion conditions, optimizing the selection process for alloys like carbon steels, stainless steels, and duplex steels. These methods excel at analyzing large, multidimensional datasets encompassing alloy composition, environmental conditions, and corrosion behavior, offering engineers predictive tools that enhance efficiency and accuracy [17] [18].

### 2.3.1. Machine Learning and Deep Learning Methods

Several ML and DL techniques are employed in material selection for corrosion resistance, each with distinct strengths:

1) **Artificial Neural Networks (ANNs):** ANNs are widely used for their ability to model nonlinear relationships in complex datasets, such as predicting corrosion rates based on alloy composition and environmental factors (e.g., pH, temperature, $CO_2$/$H_2S$ partial pressures). They consist of interconnected nodes that learn patterns through training on historical corrosion data, making them suitable for forecasting material performance in pipelines [17] [19].

2) **Support Vector Machines (SVMs):** SVMs are effective for classifying materials as suitable or unsuitable for corrosive environments, such as sour service with high $H_2S$ content. By mapping data into higher dimensional spaces, SVMs identify optimal boundaries for material performance, offering high accuracy in scenarios with limited training data [20]. They are particularly useful for binary decisions, such as selecting between stainless and duplex steels for pitting resistance [21].

3) **Decision Trees and Ensemble Methods:** Decision trees provide interpretable models by breaking down material selection into hierarchical decisions based on features like alloy microstructure or chloride exposure. Ensemble methods, such as random forests or gradient boosting, combine multiple trees to improve prediction robustness, making them ideal for handling heterogeneous pipeline

corrosion data [22] [23].

**4) Deep Learning Models:** DL models, such as convolutional neural networks (CNNs) or physics informed neural networks (PINNs), are increasingly applied to complex corrosion scenarios. CNNs can analyze microstructural images to correlate grain size or phase distribution with corrosion resistance, while PINNs integrate physical corrosion laws to enhance prediction accuracy [23] [24]. These models are particularly effective for modeling multipoint corrosion or stress corrosion cracking (SCC) in pipeline steels [24].

These methods leverage large datasets, including experimental corrosion rates, field performance records, and simulated environmental conditions, to train models that predict material behavior with high precision [21] [25].

### 2.3.2. Data Driven Prediction of Corrosion Resistance

Data driven approaches use extensive datasets to predict corrosion resistance, reducing reliance on time consuming lab tests. Alloy composition (elements like chromium, molybdenum, and nickel enhances resistance to pitting and SCC in stainless and duplex steels), environmental factors (variables such as $CO_2/H_2S$ partial pressures, temperature, pH, and chloride concentration, critical for onshore pipeline corrosion), microstructural properties (grain size, phase distribution, and inclusions, which influence localized corrosion like pitting or MIC), mechanical properties (yield strength and toughness, which affect susceptibility to SCC under pipeline stresses). AI models process these inputs to predict outcomes like corrosion rates (mm/year), pitting depth, or time to failure under specific conditions. For instance, ANNs have been used to model the corrosion rate of carbon steel pipelines in $CO_2$ rich environments, achieving prediction errors below 5% compared to experimental data [17] [19]. Similarly, SVMs and random forests can classify materials as resistant or susceptible to $H_2S$ induced cracking, guiding the selection of duplex steels for sour service [21] [22]. These predictions enable proactive material selection, optimizing for both performance and cost in onshore pipelines [23].

### 2.3.3. Case Studies

Several case studies highlight the efficacy of AI in material selection and corrosion prediction. High entropy alloys [25] applied deep learning to predict corrosion resistant compositions in high entropy alloys, identifying novel alloys with superior resistance to pitting and SCC. This approach is adaptable to pipeline steels, where AI can screen compositions for optimal chromium or molybdenum content. Corrosion rate forecasting in pipelines [17] developed an ANN model to predict internal corrosion rates in oil pipelines, using inputs like fluid composition and flow rate. The model achieved high accuracy in forecasting $CO_2$ corrosion, aiding the selection of carbon vs. stainless steels. Inhibitor efficiency modeling [19] used ANNs to predict the effect of corrosion inhibitors on $CO_2$ corrosion in carbon steels, demonstrating how AI can optimize material performance by factoring in chemical treatments. Stress corrosion cracking prediction [26] employed ANNs to predict time to failure for 304 stainless steels under SCC conditions, providing insights for selecting materials resistant to tensile stresses in pipelines. Microbiologically influenced corrosion (MIC) [27] applied AI to model MIC in offshore systems, adaptable to onshore pipelines, to predict degradation rates and select resistant materials like duplex steels. Vision based corrosion detection [28] used CNNs with micro aerial vehicles to detect corrosion on pipeline surfaces, supporting material selection by identifying early degradation patterns. Environmental assisted cracking, studies like [29] used ANNs and risk-based AI models to predict corrosion in equatorial or $H_2S/CO_2$ environments, informing material choices for harsh onshore conditions.

Implementing AI for material selection requires robust datasets, often sourced from experimental corrosion tests, field data, or simulations. Challenges include data quality (e.g., incomplete or noisy datasets) and computational complexity, particularly for DL models like CNNs or PINNs [24] [30]. Feature selection is critical, as irrelevant inputs (e.g., non-corrosive environmental factors) can reduce model accuracy. Techniques like principal component analysis (PCA) or feature importance ranking in random forests help prioritize key parameters like alloy composition or chloride exposure [22]. Additionally, hybrid approaches combining AI with physical corrosion models (e.g., NACE standards) enhance reliability by grounding predictions in established science. AI assisted material selection leverages ML and DL to revolutionize the identification of corrosion resistant steels for onshore pipelines. By combining predictive power with case study validation, these methods enable engineers to select materials that balance performance, cost, and durability, addressing the unique corrosion challenges of $CO_2$, $H_2S$, MIC, pitting, and SCC.

### 2.4. Framework Development for Corrosion Resistant Steels

Developing an AI-assisted framework for selecting corrosion resistant steels for onshore oil and gas pipelines involves a systematic approach to integrate material science, environmental data, and advanced computational techniques. The framework aims to optimize material selection by predicting performance under corrosive conditions, reducing costs, and enhancing pipeline longevity. This process encompasses defining comprehensive input parameters, training and validating AI models with robust datasets, and designing decision support systems to provide actionable recommendations for engineers. The framework leverages machine learning to bridge data gaps, improve prediction accuracy, and streamline decision making in complex, corrosive environments typical of onshore pipelines. The framework relies on a multidimensional set of input parameters to capture the factors influencing corrosion resistance and material suitability.

**Alloy Composition:** Key elements such as chromium, molybdenum, nickel, and carbon content significantly affect corrosion resistance. For example, higher chromium content enhances pitting resistance in stainless and duplex steels, quantified by the Pitting Resistance Equivalent Number [21] [25].

**Microstructure**: Grain size, phase distribution (e.g., austenite, ferrite in duplex steels), and precipitate formation influence mechanical strength and susceptibility to stress corrosion cracking (SCC) or pitting [22]. Microstructural data can be derived from metallographic analysis or simulations.

**Mechanical Properties**: Yield strength, tensile strength, and hardness are critical for ensuring pipelines withstand operational stresses while resisting corrosion induced degradation [24]. For instance, duplex steels offer high yield strength for high pressure pipelines.

**Corrosion Properties:** Metrics like corrosion rate (mm/year), PREN, and susceptibility to specific mechanisms (e.g., $CO_2$ corrosion, $H_2S$ induced cracking) are evaluated under simulated or field conditions [19]. These properties are often tested per standards like NACE TM0177 for sour service.

**Service Environment:** Environmental factors include pH, temperature, partial pressures of $CO_2$ and $H_2S$, chloride concentration, soil corrosivity, and flow conditions (e.g., velocity, turbulence). These parameters dictate corrosion severity and material performance in onshore settings [23] [31]. For example, high $CO_2$ partial pressures exacerbate sweet corrosion, necessitating corrosion resistant alloys (CRAs).

These inputs are compiled from experimental data, field measurements, and industry databases, ensuring the framework accounts for real world variability in pipeline operations [30].

### AI Model Training and Validation

AI model development for selecting corrosion resistant steels in onshore oil and gas pipelines involves selecting appropriate algorithms and training them on high quality datasets to predict material performance under corrosive conditions. Common machine learning approaches include artificial neural networks (ANNs), which excel at modeling complex, non-linear relationships between alloy composition and corrosion rates, and support vector machines (SVMs), which classify materials as suitable or unsuitable based on environmental thresholds [22]. Decision trees and random forests offer interpretable predictions, making them accessible for engineers less familiar with AI [27]. Physics informed neural networks (PINNs) enhance accuracy by incorporating physical laws, such as electrochemical kinetics and mass transport equations, reducing reliance on large datasets, a critical advantage for niche applications like onshore pipelines where data may be scarce [24]. For example, PINNs can model $CO_2$ corrosion rates by integrating Faraday's laws with environmental data. Training datasets are sourced from laboratory experiments (e.g., autoclave tests for $H_2S$ corrosion), field corrosion monitoring (e.g., weight loss coupons, electrochemical probes), and computational simulations (e.g., finite element modeling of stress corrosion cracking) [26], supplemented by public databases like NACE Corrosion Data or proprietary industry datasets. Model validation employs techniques like k-fold cross validation to ensure generalizability across diverse pipeline conditions, such as testing a model trained on carbon steel corrosion in $CO_2$ environments across varying temperatures and pH levels to confirm robustness [32]. Performance is quantified using metrics like mean squared error for regression tasks or F1 score for classification, with hyperparameter tuning (e.g., adjusting learning rates or ANN hidden layers) optimizing outcomes [33]. Data preprocessing, including handling missing values, normalizing inputs like PREN or corrosion rates, and addressing outliers, is essential for reliable predictions, while techniques like synthetic data generation or transfer learning mitigate data scarcity for rare corrosion scenarios, such as microbiologically influenced corrosion in specific soil types [34]. **Figure 3** shows AI-assisted framework for selecting corrosion resistant for an onshore oil and gas pipelines.

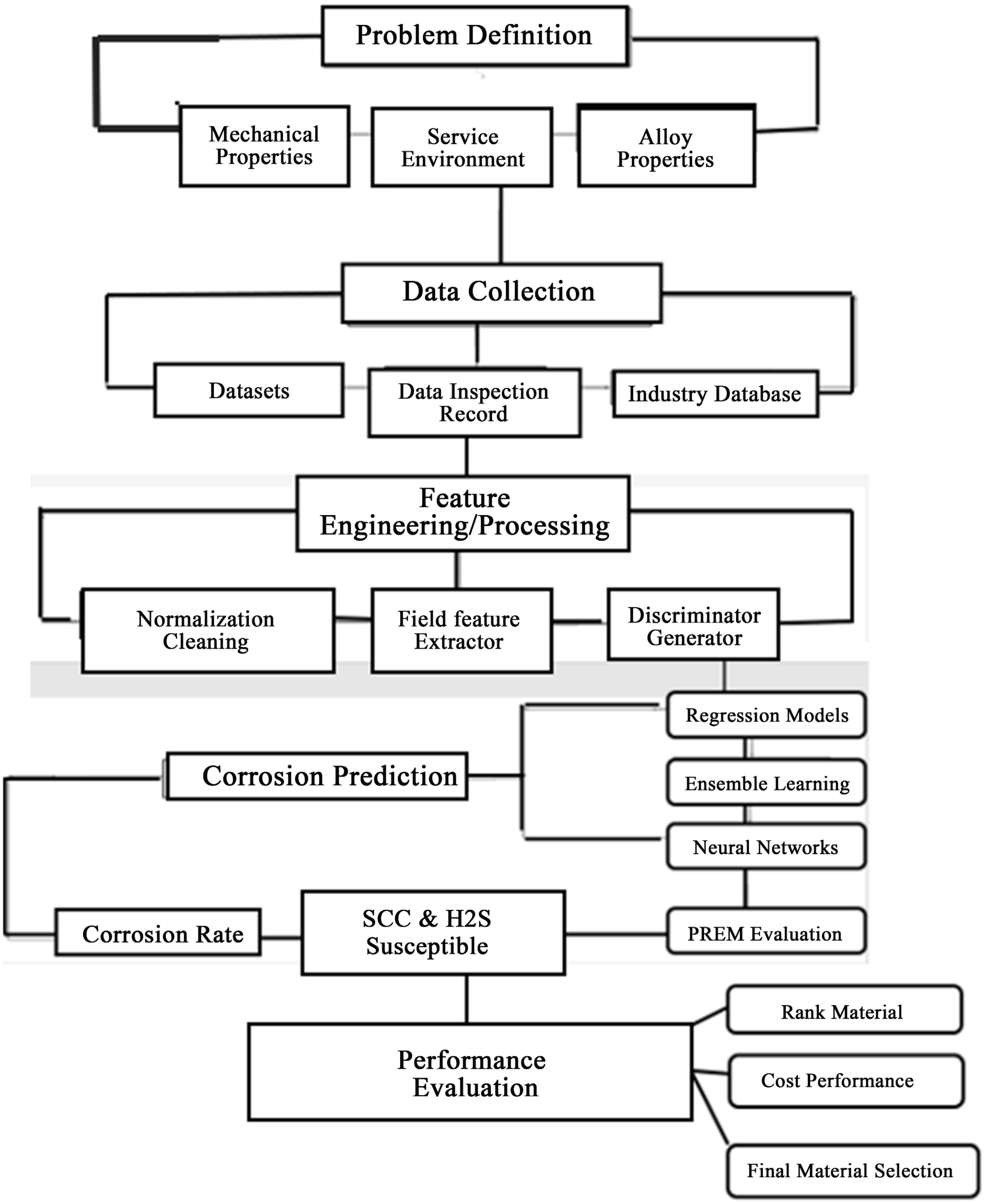

**Figure 3.** AI-assisted framework for selecting corrosion resistant for an onshore oil and gas pipelines.

### 2.5. Comparative Analysis

The integration of artificial intelligence (AI) into material selection for corrosion resistant steels in onshore oil and gas pipelines offers transformative advantages over traditional methods, but it also introduces unique challenges. This compares the strengths and weaknesses of AI driven approaches, such as machine learning (ML) and deep learning (DL), against conventional methods like performance testing and empirical modeling, while addressing methodological hurdles including data scarcity, generalizability, and interpretability.

#### 2.5.1. Strengths of AI Methods

AI techniques, such as artificial neural networks (ANNs), support vector machines (SVMs), and ensemble learning, excel in processing complex, high dimensional datasets to predict corrosion behavior and material performance with high accuracy [35]. Unlike traditional methods, which rely on time consuming laboratory tests, AI models can rapidly analyze vast datasets, identifying patterns in alloy composition, environmental conditions, and corrosion rates that may elude empirical models [36]. For instance, ANNs have been used to predict $CO_2$ corrosion rates in carbon steels with errors below 5%, significantly faster than electrochemical testing [37]. AI's scalability allows it to handle diverse pipeline conditions, from sour ($H_2S$ rich) to sweet ($CO_2$ rich) environments, enabling proactive material selection for specific onshore settings [38]. Additionally, AI frameworks can integrate multi-objective optimization, balancing cost, durability, and corrosion resistance, as demonstrated in decision support systems for duplex steel selection [39]. These systems reduce human bias and streamline decision making compared to subjective cost benefit analyses.

#### 2.5.2. Weaknesses of AI Methods

Despite their advantages, AI methods depend heavily on high quality, comprehensive datasets, which are often scarce for niche applications like onshore pipelines exposed to microbiologically influenced corrosion (MIC) or stress corrosion cracking (SCC) [40]. Unlike traditional methods, which are grounded in well-established physical principles (e.g., Faraday's laws for corrosion kinetics), AI models, particularly deep learning, can act as "black boxes", offering limited insight into the physical mechanisms driving predictions. This lack of interpretability can undermine trust among engineers accustomed to transparent empirical models like those in NACE MR0175 standards [41]. Furthermore, AI models trained on specific datasets (e.g., laboratory-based corrosion tests) may struggle to generalize to real world pipeline conditions, such as varying soil chemistries or fluctuating temperatures.

#### 2.5.3. Strengths of Conventional Methods

Traditional approaches, such as immersion testing, electrochemical analysis, and empirical models, provide reliable, physics-based insights into corrosion behavior. For example, standards like NACE MR0175 offer clear guidelines for selecting materials in $H_2S$ environments, ensuring compliance with industry regulations. These methods are well validated, with decades of field data supporting their use, such as corrosion rate predictions for carbon steels in $CO_2$ rich pipelines [13]. Their transparency allows engineers to trace the rationale behind material choices, making them suitable for critical applications where safety is paramount.

#### 2.5.4. Weaknesses of Conventional Methods

Conventional methods are often labor intensive and time consuming, requiring extensive testing to evaluate materials under simulated pipeline conditions. For instance, pitting corrosion tests for stainless steels can take weeks to yield results [42]. Empirical models, while robust, rely on historical data that may not account for novel alloys or extreme environments, leading to conservative designs or overlooked risks. Subjectivity in interpreting test results or weighing cost benefit tradeoffs can also introduce inconsistencies, particularly when selecting between carbon and duplex steels for aggressive onshore settings [43].

#### 2.5.5. Methodological Challenges for AI

Several methodological challenges impede the widespread adoption of AI in material selection for corrosion resistant steels in onshore oil and gas pipelines. Data scarcity poses a significant hurdle, as comprehensive datasets linking alloy composition, microstructure, and corrosion performance in specific environments, such as high chloride soils, are often limited. For instance, microbiologically influenced corrosion (MIC) data for onshore pipelines is frequently incomplete due to the complex interplay of microbial interactions, leading to overfitting in AI models and reduced reliability [44]. Generalizability is another critical issue, as AI models trained on controlled laboratory conditions may struggle to adapt to diverse field conditions, such as varying pH levels or $H_2S$ partial pressures encountered in onshore pipelines [38]. While transfer learning and domain adaptation techniques are being explored to enhance generalizability, these approaches demand substantial computational resources, complicating their practical implementation [45]. Additionally, the interpretability of black box models, such as deep neural networks, remains a challenge, as these models obscure the causal relationships between inputs like chromium content and outputs like corrosion rate, making it difficult to justify material selections to stakeholders. Hybrid approaches, such as physics-informed neural networks, offer a promising solution by integrating physical laws to improve interpretability, as demonstrated in models predicting stress corrosion cracking (SCC) in stainless steels [46].

## 3. Discussion

The integration of artificial intelligence (AI) into material selection for corrosion resistant steels in onshore oil and gas pipelines represents a transformative shift in corrosion management, with significant implications for both practice and theory. This section discusses major trends in AI's growing role in predictive corrosion modeling, identifies critical knowledge gaps, and explores the practical and theoretical implications of adopting AI-assisted frameworks.

### 3.1. Discussion on Major Trends

AI's role in predictive corrosion modeling is rapidly expanding, driven by advancements in machine learning (ML) and deep learning (DL) techniques such as artificial neural networks (ANNs), support vector machines (SVMs), and ensemble methods [35] [37]. These models excel at analyzing complex datasets, predicting corrosion rates, and identifying optimal material compositions for challenging environments like high chloride soils or $H_2S$-rich pipelines [25]. For example, ANNs have achieved prediction errors below 5% for $CO_2$ corrosion in carbon steels, far surpassing the speed of traditional electrochemical testing [37]. The emergence of hybrid AI-physics models, such as physics informed neural networks, further enhances predictive accuracy by incorporating corrosion kinetics, as demonstrated in studies modeling stress corrosion cracking (SCC) in stainless steels [46] [47]. Additionally, AI-driven decision support systems are streamlining material selection by optimizing tradeoffs between cost, durability, and corrosion resistance, reducing reliance on subjective cost benefit analyses [17]. This trend reflects a broader shift toward data-driven materials science, with AI enabling proactive rather than reactive corrosion management strategies.

### 3.2. Discussion Knowledge Gaps

Despite these advancements, significant knowledge gaps hinder AI's full adoption in pipeline material selection. A primary challenge is data scarcity, as comprehensive datasets linking alloy composition, microstructure, and corrosion performance in specific onshore environments (an example is high chloride soils or microbiologically influenced corrosion [MIC] settings) are limited. For instance, MIC data is often incomplete due to the complexity of microbial interactions, leading to overfitting and reduced model reliability [48]. Another gap is the lack of field validation for AI models, which are frequently trained on controlled laboratory data that may not capture the variability of real-world pipeline conditions, such as fluctuating pH or $H_2S$ partial pressures [38]. While transfer learning and domain adaptation are being explored to improve generalizability, these techniques require significant computational resources and are not yet widely validated in pipeline applications [49]. Furthermore, the interpretability of black box models like deep neural networks remains a barrier, as their opaque decision-making processes can undermine stakeholder trust compared to transparent empirical models like those in NACE MR0175 [41] [50]. These gaps highlight the need for robust datasets and field-tested AI frameworks to ensure reliability in practical applications [51].

The data generation in this strategy consists of two parts [52] the generation of input variables (environmental factors) and output variables (corrosion depth) for corrosion growth prediction. For the outliers in the corrosion dataset, multiple detection methods were used for data cleaning. For tabular data and oversampling, CTGAN, which introduces mode-specific normalization and training by sampling, is used to learn real data to generate new environmental factors. For the regression challenge, a corrosion growth model based on machine learning algorithms is applied to generate corrosion depths corresponding to the new environmental factors. The fake data are merged with the original data to form the synthetic dataset. Finally, the effectiveness of the proposed strategy is verified by analyzing the distribution and credibility of the synthetic dataset using various visualization methods and evaluation metrics. This technique can provide data support for future corrosion growth modeling, and this method can also be applied to data generation for regression problems in other fields.

## 3.3. Implications for Practice

The adoption of AI-assisted material selection frameworks promises more reliable and cost-effective solutions for onshore pipeline corrosion management. By rapidly analyzing vast datasets, AI models can identify corrosion resistant steels (e.g., duplex or stainless steels) tailored to specific environmental conditions, reducing the risk of pipeline failures due to $CO_2$, $H_2S$, or MIC [27] [37]. AI-driven decision-support systems can rank materials based on multi-objective optimization, balancing corrosion resistance with lifecycle costs, which is particularly valuable for resource-constrained operators [21]. These systems also minimize human bias inherent in traditional methods, such as subjective interpretations of test results. Moreover, hybrid AI-physics models improve prediction reliability by integrating physical principles, as seen in $CO_2$ corrosion models achieving 10% higher accuracy than purely data-driven approaches [24]. In practice, these advancements could reduce maintenance costs and downtime, enhancing pipeline safety and longevity. However, practitioners must address data scarcity and model validation to ensure AI recommendations align with field performance and regulatory standards.

Materials science can also boost machine learning model explainability and performance in materials problems, as we have seen in the [53]. Physics informed machine learning is a blooming research field [54], instance according [55] topics that are related are general informed machine learning and theory guided data science Uncertainty quantification is another interesting field that can help ensure trust in ML model.

## 4. Conclusions

The integration of artificial intelligence (AI) into material selection is not limited to corrosion-resistant steels in onshore oil and gas pipelines marking a significant advancement in corrosion management Materials science can also boost machine learning model explainability and performance in materials problems, as we have seen in the [53] instance According [55] Physics informed machine learning is a blooming research field [54], topics that are related are general informed machine learning and theory guided data science Uncertainty quantification is another interesting field that can help ensure trust in ML model. Key insights from this study highlight that AI, through machine learning techniques like artificial neural networks and hybrid AI-physics models, enhances the reliability of selecting materials such as duplex and stainless steels by accurately predicting corrosion [47]. These AI-driven frameworks outperform traditional methods by rapidly analyzing vast datasets, optimizing trade-offs between cost and durability, and reducing human bias inherent in empirical models [21] [37]. This reliability is critical for ensuring pipeline safety, preventing failures due to $CO_2$, $H_2S$, or microbiologically influenced corrosion (MIC), and achieving significant cost savings by minimizing maintenance and downtime [27]. The importance of these advancements cannot be overstated, as corrosion-related pipeline failures pose severe safety risks and economic losses in the oil and gas industry [38]. Looking ahead, future research should focus on three key directions:

1) Developing larger, comprehensive material databases to address data scarcity, particularly for niche conditions like high-chloride soils or MIC.

2) Advancing hybrid AI and physics-based models to improve interpretability and prediction accuracy, as demonstrated in recent $CO_2$ corrosion studies.

3) Integrating AI with real-time monitoring systems to enable dynamic material performance assessments, leveraging technologies like IoT sensors for continuous data collection.

These efforts will further solidify AI's role in revolutionizing material selection, enhancing pipeline integrity, and advancing materials informatics for corrosion-resistant alloys.

## Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

## References

[1] Papadimitriou, I., Gialampoukidis, I., Vrochidis, S. and Kompatsiaris, I. (2024) AI Methods in Materials Design, Discovery and Manufacturing: A Review. *Computational Materials Science*, **235**, Article ID: 112793. https://doi.org/10.1016/j.commatsci.2024.112793

[2] Merayo, D., Rodríguez-Prieto, A. and Camacho, A.M. (2019) Comparative Analysis of Artificial Intelligence Techniques for Material Selection Applied to Manufacturing in Industry 4.0. *Procedia Manufacturing*, **41**, 42-49. https://doi.org/10.1016/j.promfg.2019.07.027

[3] Ebili, F., Oterkus, S. and Oterkus, E. (2025) A Detailed Structural Review of Onshore and Offshore Pipelines Containing Defects. *Sustainable Marine Structures*, **7**, 177-208. https://doi.org/10.36956/sms.v7i3.2320

[4] Ossai, C.I. (2019) A Data-Driven Machine Learning Approach for Corrosion Risk Assessment—A Comparative Study. *Big Data and Cognitive Computing*, **3**, Article 28. https://doi.org/10.3390/bdcc3020028

[5] Nešić, S. (2007) Key Issues Related to Modelling of Internal Corrosion of Oil and Gas Pipelines—A Review. *Corrosion Science*, **49**, 4308-4338. https://doi.org/10.1016/j.corsci.2007.06.006

[6] Kermani, M.B. and Morshed, A. (2003) Carbon Dioxide Corrosion in Oil and Gas Production—A Compendium. *Corrosion*, **59**, 659-683. https://doi.org/10.5006/1.3277596

[7] Ossai, C.I., Boswell, B. and Davies, I.J. (2015) Pipeline Failures in Corrosive Environments—A Conceptual Analysis of Trends and Effects. *Engineering Failure Analysis*, **53**, 36-58. https://doi.org/10.1016/j.engfailanal.2015.03.004

[8] Chen, J., Althaus, S.M., Liu, H. and Sun, Q. (2019) Shale Gas Transport in Rock Matrix: Diffusion in the Presence of Surface Adsorption and Capillary Condensation. *Journal of Natural Gas Science and Engineering*, **66**, 18-25. https://doi.org/10.1016/j.jngse.2019.03.015

[9] Popov, B.N. (2015) Evaluation of Corrosion. In: Popov, B.N., Ed., *Corrosion Engineering*, Elsevier, 1-28. https://doi.org/10.1016/b978-0-444-62722-3.00001-x

[10] Kaddor, C. and Steinbüchel, A. (2011) Implications of Various Phosphoenolpyruvate-Carbohydrate Phosphotransferase System Mutations on Glycerol Utilization and Poly(3-Hydroxybutyrate) Accumulation in *Ralstonia eutropha* H16. *AMB Express*, **1**, Article No. 16. https://doi.org/10.1186/2191-0855-1-16

[11] Ashby, M.F., Bréchet, Y.J.M., Cebon, D. and Salvo, L. (2004) Selection Strategies for Materials and Processes. *Materials & Design*, **25**, 51-67. https://doi.org/10.1016/s0261-3069(03)00159-6

[12] Reda, A., Shahin, M.A. and Montague, P. (2025) Review of Material Selection for Corrosion-Resistant Alloy Pipelines. *Engineered Science*, **33**, Article 1373. https://doi.org/10.30919/es1373

[13] Wei, R.P. and Gao, M. (1991) Technical Note: Distribution of Initial Current between Bare and Filmed Surfaces: What Is Being Measured in a Scratched Electrode Test? *Corrosion*, **47**, 948-951. https://doi.org/10.5006/1.3585207

[14] Caleyo, F., Velázquez, J.C., Valor, A. and Hallen, J.M. (2009) Markov Chain Modelling of Pitting Corrosion in Underground Pipelines. *Corrosion Science*, **51**, 2197-2207. https://doi.org/10.1016/j.corsci.2009.06.014

[15] Jahan, A., Ismail, M.Y., Sapuan, S.M. and Mustapha, F. (2010) Material Screening and Choosing Methods—A Review. *Materials & Design*, **31**, 696-705. https://doi.org/10.1016/j.matdes.2009.08.013

[16] Do, H., Rothermel, G. and Kinneer, A. (2006) Prioritizing Junit Test Cases: An Empirical Assessment and Cost-Benefits Analysis. *Empirical Software Engineering*, **11**, 33-70. https://doi.org/10.1007/s10664-006-5965-8

[17] Badings, T.S. and van Putten, D.S. (2020) Data Validation and Reconciliation for Error Correction and Gross Error Detection in Multiphase Allocation Systems. *Journal of Petroleum Science and Engineering*, **195**, Article ID: 107567. https://doi.org/10.1016/j.petrol.2020.107567

[18] Lee, S.Y., Byeon, S., Kim, H.S., Jin, H. and Lee, S. (2021) Deep Learning-Based Phase Prediction of High-Entropy Alloys: Optimization, Generation, and Explanation. *Materials & Design*, **197**, Article ID: 109260. https://doi.org/10.1016/j.matdes.2020.109260

[19] Hernández, S., Nešić, S., Weckman, G. and Ghai, V. (2006) Use of Artificial Neural Networks for Predicting Crude Oil Effect on Carbon Dioxide Corrosion of Carbon Steels. *Corrosion*, **62**, 467-482. https://doi.org/10.5006/1.3279905

[20] Ben Seghier, M.E.A., Höche, D. and Zheludkevich, M. (2022) Prediction of the Internal Corrosion Rate for Oil and Gas Pipeline: Implementation of Ensemble Learning Techniques. *Journal of Natural Gas Science and Engineering*, **99**, Article ID: 104425.

https://doi.org/10.1016/j.jngse.2022.104425

[21] Soori, M., Jough, F.K.G., Dastres, R. and Arezoo, B. (2026) AI-Based Decision Support Systems in Industry 4.0, a Review. *Journal of Economy and Technology*, **4**, 206-225. https://doi.org/10.1016/j.ject.2024.08.005

[22] Liu, L. (2024) Machine Learning-Driven Corrosion Detection and Classification in Pipelines. Ph.D. Thesis, University of Wales Trinity Saint David.

[23] Lu, H., Peng, H., Xu, Z., Qin, G., Azimi, M., Matthews, J.C., *et al.* (2023) Theory and Machine Learning Modeling for Burst Pressure Estimation of Pipeline with Multipoint Corrosion. *Journal of Pipeline Systems Engineering and Practice*, **14**, 04023022. https://doi.org/10.1061/jpsea2.pseng-1481

[24] Bansal, P., Zheng, Z., Shao, C., Li, J., Banu, M., Carlson, B.E., *et al.* (2022) Physics-informed Machine Learning Assisted Uncertainty Quantification for the Corrosion of Dissimilar Material Joints. *Reliability Engineering & System Safety*, **227**, Article ID: 108711. https://doi.org/10.1016/j.ress.2022.108711

[25] Wen, C., Zhang, Y., Wang, C., Xue, D., Bai, Y., Antonov, S., *et al.* (2019) Machine Learning Assisted Design of High Entropy Alloys with Desired Property. *Acta Materialia*, **170**, 109-117. https://doi.org/10.1016/j.actamat.2019.03.010

[26] Lajevardi, S.A., Shahrabi, T., Baigi, V. and Shafiei M, A. (2009) Prediction of Time to Failure in Stress Corrosion Cracking of 304 Stainless Steel in Aqueous Chloride Solution by Artificial Neural Network. *Protection of Metals and Physical Chemistry of Surfaces*, **45**, 610-615. https://doi.org/10.1134/s2070205109050207

[27] Adumene, S., Khan, F., Adedigba, S., Zendehboudi, S. and Shiri, H. (2021) Dynamic Risk Analysis of Marine and Offshore Systems Suffering Microbial Induced Stochastic Degradation. *Reliability Engineering & System Safety*, **207**, Article ID: 107388. https://doi.org/10.1016/j.ress.2020.107388

[28] Ortiz, A., Bonnin-Pascual, F., Garcia-Fidalgo, E. and Company-Corcoles, J. (2016) Vision-Based Corrosion Detection Assisted by a Micro-Aerial Vehicle in a Vessel Inspection Application. *Sensors*, **16**, Article 2118. https://doi.org/10.3390/s16122118

[29] Bertolini, L. and Redaelli, E. (2009) Throwing Power of Cathodic Prevention Applied by Means of Sacrificial Anodes to Partially Submerged Marine Reinforced Concrete Piles: Results of Numerical Simulations. *Corrosion Science*, **51**, 2218-2230. https://doi.org/10.1016/j.corsci.2009.06.012

[30] Nasser, A.M.M., Montasir, O.A., Wan Abdullah Zawawi, N.A. and Alsubal, S. (2020) A Review on Oil and Gas Pipelines Corrosion Growth Rate Modelling Incorporating Artificial Intelligence Approach. *IOP Conference Series: Earth and Environmental Science*, **476**, Article ID: 012024. https://doi.org/10.1088/1755-1315/476/1/012024

[31] Goswami, L., Deka, M.K. and Roy, M. (2023) Artificial Intelligence in Material Engineering: A Review on Applications of Artificial Intelligence in Material Engineering. *Advanced Engineering Materials*, **25**, 2300104. https://doi.org/10.1002/adem.202300104

[32] Blake, W.K. and Gershfeld, J.L. (1989) The Aeroacoustics of Trailing Edges. In: Gad-el-Hak, M., Ed., *Frontiers in Experimental Fluid Mechanics*, Springer, 457-532. https://doi.org/10.1007/978-3-642-83831-6_10

[33] Singh, B. and Krishnathasan, K. (2009) Pragmatic Effects of Flow on Corrosion Prediction. *CORROSION* 2009, Atlanta, 22-26 March 2009, 1-29. https://doi.org/10.5006/c2009-09275

[34] Naveen Prasad, B.S., Meka, U., Rajasekaran, R. and Banerjee, S. (2025) Machine Learning Techniques in Water Treatment. In: Namdeti, R. and Abuda Joaquin, A.,

Eds., *Machine Learning in Water Treatment*, Wiley, 345-411.

[35] Zhou, Q., Chen, X. and Wang, J. (2025) Machine Learning Assisted Material Discovery: A Small Data Approach. *Accounts of Materials Research*, **6**, 685-694. https://doi.org/10.1021/accountsmr.1c00236

[36] Morgan, D. and Jacobs, R. (2020) Opportunities and Challenges for Machine Learning in Materials Science. *Annual Review of Materials Research*, **50**, 71-103. https://doi.org/10.1146/annurev-matsci-070218-010015

[37] Al-Sabaeei, A.M., Alhussian, H., Abdulkadir, S.J. and Jagadeesh, A. (2023) Prediction of Oil and Gas Pipeline Failures through Machine Learning Approaches: A Systematic Review. *Energy Reports*, **10**, 1313-1338. https://doi.org/10.1016/j.egyr.2023.08.009

[38] Askari, M., Aliofkhazraei, M. and Afroukhteh, S. (2019) A Comprehensive Review on Internal Corrosion and Cracking of Oil and Gas Pipelines. *Journal of Natural Gas Science and Engineering*, **71**, Article ID: 102971. https://doi.org/10.1016/j.jngse.2019.102971

[39] Bhandari, J., Khan, F., Abbassi, R., Garaniya, V. and Ojeda, R. (2015) Modelling of Pitting Corrosion in Marine and Offshore Steel Structures—A Technical Review. *Journal of Loss Prevention in the Process Industries*, **37**, 39-62. https://doi.org/10.1016/j.jlp.2015.06.008

[40] Shaik, N.B., Jongkittinarukorn, K., Benjapolakul, W. and Bingi, K. (2024) A Novel Neural Network-Based Framework to Estimate Oil and Gas Pipelines Life with Missing Input Parameters. *Scientific Reports*, **14**, Article No. 4511. https://doi.org/10.1038/s41598-024-54964-3

[41] Lu, H., Xi, D., Xiang, Y., Su, Z. and Cheng, Y.F. (2025) Vehicle-Canine Collaboration for Urban Pipeline Methane Leak Detection. *Nature Cities*, **2**, 336-343. https://doi.org/10.1038/s44284-024-00183-w

[42] Mokaberi, A., Derakhshandeh-Haghighi, R. and Abbaszadeh, Y. (2015) Fatigue Fracture Analysis of Gas Turbine Compressor Blades. *Engineering Failure Analysis*, **58**, 1-7. https://doi.org/10.1016/j.engfailanal.2015.08.026

[43] Sur, D., Gupta, A., Dubey, S. and Kumar, A. (2025) Properties of Materials and Selection Criteria. *Chemical Engineering Essentials* 2: *Advanced Processes*, *Materials*, *and Sustainability*, 79-107.

[44] Yazdi, M. (2022) Management of Offshore Structures under Microbiologically Influenced Corrosion (MIC). Master's Thesis, Memorial University of Newfoundland.

[45] Canonaco, G., Roveri, M., Alippi, C., Podenzani, F., Bennardo, A., Conti, M., *et al.* (2021) A Transfer-Learning Approach for Corrosion Prediction in Pipeline Infrastructures. *Applied Intelligence*, **52**, 7622-7637. https://doi.org/10.1007/s10489-021-02771-y

[46] May, Z., Alam, M.K., Nayan, N.A., Rahman, N.A.A. and Mahmud, M.S. (2021) Acoustic Emission Corrosion Feature Extraction and Severity Prediction Using Hybrid Wavelet Packet Transform and Linear Support Vector Classifier. *PLOS ONE*, **16**, e0261040. https://doi.org/10.1371/journal.pone.0261040

[47] Qu, Z., Zou, X., Xiong, G., Yue, X. and Zhang, L. (2025) Hybrid Intelligent Model for Predicting Corrosion Rate of Carbon Steel in $Co_2$ Environments. *Materials and Corrosion*, **76**, 1319-1326. https://doi.org/10.1002/maco.202514840

[48] Hussain, A.M., Sanoussi, A.A. and Hussain, H.A.M. (2010) Pollution of Drinking Water Transported by Corroded Metallic Pipelines. *WIT Transactions on Ecology and the Environment*, **135**, 61-69. https://doi.org/10.2495/wp100061

[49] Zhang, L. and Gao, X. (2024) Transfer Adaptation Learning: A Decade Survey. *IEEE Transactions on Neural Networks and Learning Systems*, **35**, 23-44. https://doi.org/10.1109/tnnls.2022.3183326

[50] Esna-Ashari, M. (2025) Beyond the Black Box: A Review of Quantitative Metrics for Neural Network Interpretability and Their Practical Implications. *International Journal of Sustainable Applied Science and Engineering*, **2**, 1-24. https://bgsiran.ir/journal/ojs-3.1.1-4/index.php/IJSASE/article/view/133

[51] Ma, H., Geng, M., Wang, F., Zheng, W., Ai, Y. and Zhang, W. (2024) Data Augmentation of a Corrosion Dataset for Defect Growth Prediction of Pipelines Using Conditional Tabular Generative Adversarial Networks. *Materials*, **17**, Article 1142. https://doi.org/10.3390/ma17051142

[52] Schütt, K.T., Sauceda, H.E., Kindermans, P., Tkatchenko, A. and Müller, K. (2018) SchNet—A Deep Learning Architecture for Molecules and Materials. *The Journal of Chemical Physics*, **148**, Article ID: 241722. https://doi.org/10.1063/1.5019779

[53] Karniadakis, G.E., Kevrekidis, I.G., Lu, L., Perdikaris, P., Wang, S. and Yang, L. (2021) Physics-Informed Machine Learning. *Nature Reviews Physics*, **3**, 422-440. https://doi.org/10.1038/s42254-021-00314-5

[54] Zhong, X., Gallagher, B., Liu, S., Kailkhura, B., Hiszpanski, A. and Han, T.Y. (2022) Explainable Machine Learning in Materials Science. *npj Computational Materials*, **8**, Article No. 204. https://doi.org/10.1038/s41524-022-00884-7

[55] Meng, C., Griesemer, S., Cao, D., Seo, S. and Liu, Y. (2025) When Physics Meets Machine Learning: A Survey of Physics-Informed Machine Learning. *Machine Learning for Computational Science and Engineering*, **1**, Article No. 20. https://doi.org/10.1007/s44379-025-00016-0