COMPUTOLOGY: Journal of Applied Computer Science and Intelligent Technologies
Volume 3, Issue 1, Jan-Jun 2023, pp. 23-41
DOI: 10.17492/computology.v3i1.2303
https://www.journalpressindia.com/cjacsit

A Theoretical Study on Implementing Machine Learning for Safety and Security of Oil and Gas Pipelines

Surya Krishna K. A. and Aditya Raj***

ABSTRACT

Concerning the safety, security and integrity of the working conditions of the oil and gas pipeline operators, there are several challenges including identification and prevention of the leakage points, corrosion of pipes, quality check of the materials and other major defects. The oil and gas pipelines form the backbone of the concerned industry as they ensure the safe, reliable and efficient transportation of hydrocarbons from the point of extraction or refining to the point of consumption or end users. The traditional pipeline investigation method required manual intervention and was labor-intensive work. However, with the advancement of technology and advanced computational algorithms, the safety of the pipelines is ensured with the help of data and images captured from sensors, drones and other sources. This paper presents a survey of state-of-the-art machine-learning techniques and algorithms developed for pipeline monitoring. The paper also proposes a framework for incorporating machine learning techniques to ensure oil and gas pipelines' safety, efficiency, and sustainability. This study would help the operators, engineers and potential researchers in this field to identify the way to implement machine learning algorithms to identify potential problems and take corrective actions before they become fatal.

Keywords: Machine learning; Oil and gas pipeline; Sustainability; Artificial Intelligence; Safety.

1.0 Introduction

Oil and gas are critical sources of energy that play a crucial role in modern society as they are the primary sources of energy for transportation, heating, and electricity generation. Oil and gas not only power modern economies and support our daily lives but are also the main drivers of economic development and job creation. Many industries rely on oil and gas as a basic material to produce products such as plastics, synthetic fibers, sterilizing medical equipment, synthetic rubber, and other materials. Access to reliable sources of hydrocarbons and gas is also critical for national security. Countries that are heavily dependent on oil and gas imports can be vulnerable to supply disruptions and price increases, which can have severe economic and geopolitical consequences. While there is growing interest in the way of acquiring renewable energy, petroleum will continue to play an important role in meeting global energy demands. As such, it is essential to ensure the safety and efficient production, consumption and transportation of petroleum resources, and to continue investing in research and development to improve their environmental performance and sustainability. Pipelines are the crucial components of the oil and gas industry in India, facilitating the conveyance of crude oil and natural gas across large areas with significant economic and strategic benefits.

**School of Engineering, University of Petroleum and Energy Studies, Dehradun, Uttarakhand, India (E-mail: suriyakrishna072001@gmail.com)*
***Corresponding author; School of Computer Science, University of Petroleum and Energy Studies, Dehradun, Uttarakhand, India (E-mail: raj05aditya@gmail.com)*

According to the Ministry of Petroleum and Natural Gas, India's pipeline network stretches over 16,000 kilometers, transporting natural gas, crude oil, and petroleum products from refineries and ports to various demand centers across the country. The pipeline network in India is managed and administered by companies such as Indian Oil Corporation (IOC), Gas Authority of India Limited (GAIL), and Oil and Natural Gas Corporation (ONGC), among others. In addition to cost-effectiveness, pipelines also provide a reliable and secure means of transportation for crude oil and natural gas within parts of India. According to the Petroleum Planning and Analysis Cell (PPAC), pipelines accounted for 44% of the total crude oil and petroleum products transported in India in 2020-21, with a total throughput of 78.2 million metric tons.

In terms of natural gas, pipelines accounted for 55% of the total natural gas transportation in India in 2020-21, with a total throughput of 54.5 billion cubic meters. Pipelines also play a critical role in ensuring energy security and independence for India. The country is significantly dependent on crude oil and natural gas imports to satisfy its energy demand, and pipelines serve as a means of transporting these essential resources from ports and refineries to various demand centers across the country. According to the PPAC-2021, India's crude oil import dependency stood at 85.8% in 2020-21, with the country importing 214.4 million metric tons of crude oil. In terms of natural gas, India's import dependency stood at 52.1% in 2020-21, with the country importing 39.3 billion cubic meters of natural gas. Beyond their significance in transporting oil and gas, pipelines also support a significant number of jobs in India. In a report made by the Indian Brand Equity Foundation (IBEF) in 2021, the oil and gas industry in India employs over 1 million people, with pipeline transportation playing a crucial role in job creation. Pipeline transportation in India supports indirect employment in manufacturing and engineering.

However, the network faces challenges like inadequate infrastructure, limited investment, and regulatory impediments. The National Institute of Public Finance and Policy (NIPFP) reports under-utilization of pipelines, highlighting the need for increased investment and regulatory reforms. Oil and gas pipelines are crucial for global energy infrastructure, but traditional inspection methods are time-consuming, costly, and limited. Machine learning has emerged as a promising technology for enhancing safety and efficacy in pipelines. By analyzing data from sensors and cameras, machine learning can predict and prevent issues, minimizing downtime, reducing maintenance costs, and improving safety outcomes. The paper presents a review study of the existing literature on machine learning in oil and gas pipelines. This includes a review of the different types of machine learning algorithms that have been applied to pipeline monitoring, such as supervised learning, unsupervised learning, and reinforcement learning. Also, the paper reviews the Machine learning role in improving the safety, efficiency, and sustainability of oil and gas pipelines. ML algorithms can help operators identify potential problems and take corrective action before they become major safety hazards.

The proposed work contributes to understanding the application of ML in the Oil and Gas Industry (OGI) in the following way:

- To explore the potential ML applications in oil and gas pipelines: This paper investigates and uncovers the various potential uses and applications of ML techniques within the oil and gas pipeline industry.
- To examine the different ML algorithms and techniques suitable for pipeline monitoring, fault detection, and predictive maintenance: The study delves into the analysis of diverse ML algorithms and techniques.

- To analyze the benefits and challenges associated with implementing ML in the oil and gas industry:

The paper thoroughly investigates the pros and cons of implementing ML methodologies in the oil and gas sector.

2.0 Background

2.1 Machine learning

In machine learning, the computing and processing devices are made to learn and modify themselves to achieve higher performance output. The adaptability of machine learning models finds its application in the fields of image recognition, textual analysis, sentiment analysis, video games, and robotics. The three broad methods of learning used by machines are supervised, unsupervised, and semi-supervised learning. Supervised learning involves the use of labeled data information to validate the performance of the model in terms of accuracy [1]. The validation helps in improving the performance of the model. Some of the supervised learning models are k-Means, Decision Tree classifiers, Support Vector Machines, and Regression models. Unsupervised learning involves the identification of underlying relations between the data points and grouping the points based on the similarity between them. Clustering, dimensionality reduction techniques and associative rule mining are examples of unsupervised learning.

ML has become increasingly essential in the oil and gas industry, particularly in the context of pipelines. It can enhance safety, and efficiency, and make more informed decisions about pipeline maintenance, repair, and operation. ML algorithms commonly implemented in OGI tasks include Support Vector Machine (SVM), Artificial Neural Networks (ANN), Deep Learning (DL), and Genetic Algorithms (GA). In summation, machine learning has become an essential instrument in the oil and gas industry, particularly in the context of pipelines. By leveraging the power of data analytics and machine learning, operators can enhance safety, and efficiency, and make more informed decisions about pipeline maintenance, repair, and operation.

2.2 ML in the oil and gas industry

Machine learning (ML) has acquired significant significance in the oil and gas industry, particularly in pipelines. ML algorithms can analyze massive quantities of data from sensors and other sources in real-time, identifying potential safety hazards like leaks, corrosion, and pressure anomalies. This enables operators to take corrective action before safety incidents materialize. ML can also improve the efficacy of pipeline operations by optimizing flow rates, reducing energy consumption, and enhancing maintenance scheduling. ML can also predict when maintenance is required, preventing unexpected downtime, and enhancing safety. ML algorithms can also provide valuable insights into pipeline performance and trends, enabling personnel to make more informed decisions about maintenance, repair, and operation. ML has significant potential in the oil and gas industry (OGI), particularly in the analysis and interpretation of data. Common ML algorithms used in OGI tasks include Support Vector Machine (SVM), Artificial Neural Networks (ANN), Deep Learning (DL), and Genetic Algorithms (GA). The proper functioning and maintenance of oil and gas pipelines are crucial for assuring the uninterrupted movement of energy resources. By leveraging the power of ML algorithms, operators can detect anomalies, predict failures, and make informed decisions to enhance overall pipeline performance.

3.0 Literature Review

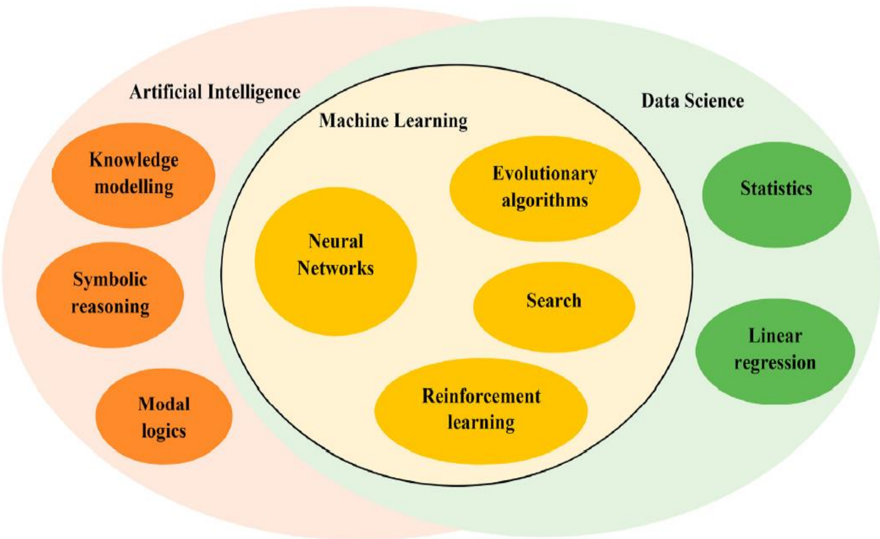
This section presents a comprehensive review of extant literature on the application of machine learning algorithms in the oil and gas industry, with emphasis on pipeline systems. The review emphasizes the various ML algorithms and techniques used for duties such as leak detection, corrosion monitoring, predictive maintenance, and anomaly detection. It also discusses the limitations and potential challenges encountered when employing ML in real-world pipeline scenarios. The methodology adopted for this research involves a detailed understanding of concepts of machine learning and its relevance to oil and gas pipelines. The research will predominantly be based on a review of scholastic articles, industry reports, and case studies. The information obtained will be synthesized and organized to provide a comprehensive understanding of the topic.

This section delves into the specific machine-learning techniques that can be implemented to enhance pipeline operations. It investigates algorithms such as decision trees, support vector machines, neural networks, and clustering algorithms, and discusses their suitability for diverse pipeline-related tasks. The focus will be on how these techniques can enhance breach detection, predict maintenance requirements, optimize flow rates, and improve operational efficiency. Implementing machine learning in oil and gas pipelines offers numerous benefits, such as improved safety, reduced maintenance costs, enhanced operational efficiency, and superior decision-making. However, there are challenges that need to be addressed, including data integrity and availability, algorithm transparency, interpretability, and regulatory compliance.

3.1 Application of machine learning in the oil and gas industry

In their paper [2], authors provided a thorough review of the role of data science and ML in various segments of the petroleum engineering and geosciences disciplines, including unconventional reservoirs. A comparative study of various techniques that have been applied in the oil and gas industry was presented by them. The issue of integration of ML and big data management techniques to address the problems of the oil and gas sector was also discussed and was brought under the purview of analytical or numerical solutions. It was noted that ML tools could integrate all details in log data and target data without being constrained by data or power processing requirements. As the oil and gas industry continues to generate substantial data in its routine activities, machine learning and big data management techniques have become indispensable for ensuring a more efficient industry.

Figure 1: Logic Diagram Showing Concepts of AI, ML, and DL [5]

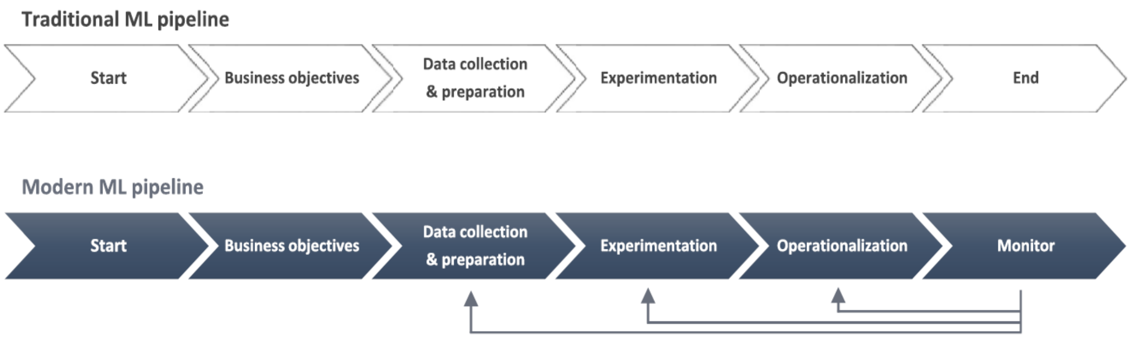


The immense potential of AI in petroleum engineering, particularly in prediction, categorization, and grouping, was underscored by the authors. The involvement of natural materials and processes in petroleum engineering and geosciences renders traditional numerical methods impractical and cumbersome. The sheer volume of data generated daily exceeds the capabilities of evaluation using traditional methods. However, the AI algorithms, along with IoT and real-time data transmission in drilling and production facilities, may help remove the risks and enhance the output [3]. Figure 1 shows the logic diagram of concepts of artificial intelligence (AI), deep learning (DL) and machine learning (ML). In their paper [4] authors performed an all-inclusive study on how AI and machine learning techniques can be applied in the oil and gas industry. The potential benefits of these technologies were emphasized, including the efficient dispensation and interpretation of large quantities of data and the performance of numerical calculations. The research on AI and ML in the upstream OGI was summarized, with a focus on their applications and limitations. It was suggested by the authors that these technologies could lead to a reduction in risk and maintenance costs, thereby simplifying decision-making processes. Due to its rapid data analysis capabilities, machine learning is expected to find increasing usage in the oil and gas sector, positioning it as a valuable resource for businesses in the future [4].

Transfer learning acquires the knowledge while training in one source task and uses this knowledge in target task environments, having different data distributions, to infer the desired output. Tan et al. showed in their work that transfer learning could be a promising approach in OGI where data are rarely available and costly [6].

Another aspect absent in the industry pertains to the utilization of Continuous Integration/Continuous Deployment (CICD) practices in ML applications, as shown in Figure 2. A contemporary CICD approach should encompass a dependable and reproducible ML pipeline, incorporating monitoring, model lineage, and version control mechanisms. This becomes especially advantageous in dealing with concept drift, wherein the performance of a predictive model declines over time due to changes in data and previously modeled input-output relationships [7].

Figure 2: A Modern ML Pipeline with a CICD Workflow



3.2 Application of machine learning in oil and gas pipeline

With the objective of using Magnetic Flux Leakage (MFL) sensors to localize the defects, Mohamed et al. studied the utility of Artificial Neural Networks (ANN) in estimating the depth at which defect has occurred in the pipeline [9]. The huge amount of raw data required by the MFL sensors posed a huge challenge in pipeline probing. The proposed work used the discriminant features, derived from unprocessed data, to train the neural networks. The proposed approach demonstrated superior performance with high accuracy within ±10% and ±15% error-tolerance ranges as compared to existing pipeline inspection techniques like GE and ROSEN. This study makes a valuable contribution to the field of oil and gas pipeline inspection and underscores the potential of machine learning approaches in handling large data complexities in the industry [9].

In [10], the issue of leakage in oil and gas pipelines using anomaly detection models of machine learning has been addressed. . Among the five anomaly detection algorithms, the Support Vector Machine outperformed with an accuracy of 97.4%. The importance of cost-effective, faster, and simpler engineering technologies for detecting pipeline leakage was highlighted in this paper. A dataset was utilized to create predictive models using various ML algorithms, and the outperforming model was selected. The proposed model exhibited excellent performance when applied to industrial data, effectively detecting pipeline leakage and preventing environmental damage and harm to industrial companies. Future work will involve the incorporation of deep learning techniques and real data from oil and gas companies to construct the models [10].

An extensive use of various analytical modeling techniques in the oil and gas sector to support data-driven decision-making is discussed [11]. Recent developments in the field of AI and ML techniques across the entire oil and gas value chain, from crude oil exploration to product distribution, were examined in this paper. The study categorized AI and ML algorithms in the oil and gas industry into nine categories, with SVM and ANN identified as the most used techniques. These methods find application in the subfields of subsurface drilling, exploring the oil fields, drilling operations, production and reservoir management, and operations related to the transportation of oil and gas [11]. Despite the progress made, challenges persist in quantifying uncertainty, minimizing risk, maximizing profit, and ensuring efficiency in decision-making processes. To promote data-driven decision-making and enhance enterprise value, oil and gas companies should make use of under-utilized data.

A detail on various factors that would affect the decision to adopt ML techniques and study the technical consequences of AI on OGI has been discussed in [12] . The study assessed the challenges and opportunities of implementing AI in the energy sector, including evaluating ML development platforms and network architecture. It explored the diverse applications of ML in exploration, drilling, reservoir management, production, transportation, and refining processes. Additionally, AI implementation trends, algorithms, data availability, and non-technical barriers were analyzed, with an emphasis on AI's ability to reduce risks [12].

A methodology based on a multi-agent system (MAS) architecture to prevent and manage hydrocarbon pipeline damages has been proposed [13]. The authors developed a system utilizing a MAS algorithm for monitoring and preventing pipeline hazards. Within this system, the ML Agent detects anomalous pressure conditions, alerts the Mobile Agent, and identifies metallic objects in the conduit. Furthermore, the metallic objects are detected by the sensor agents and notified by the Master Agent, while the Protective Agent provides immediate protection. The cooperative nature of this system indicates its suitability for addressing pipeline damages.

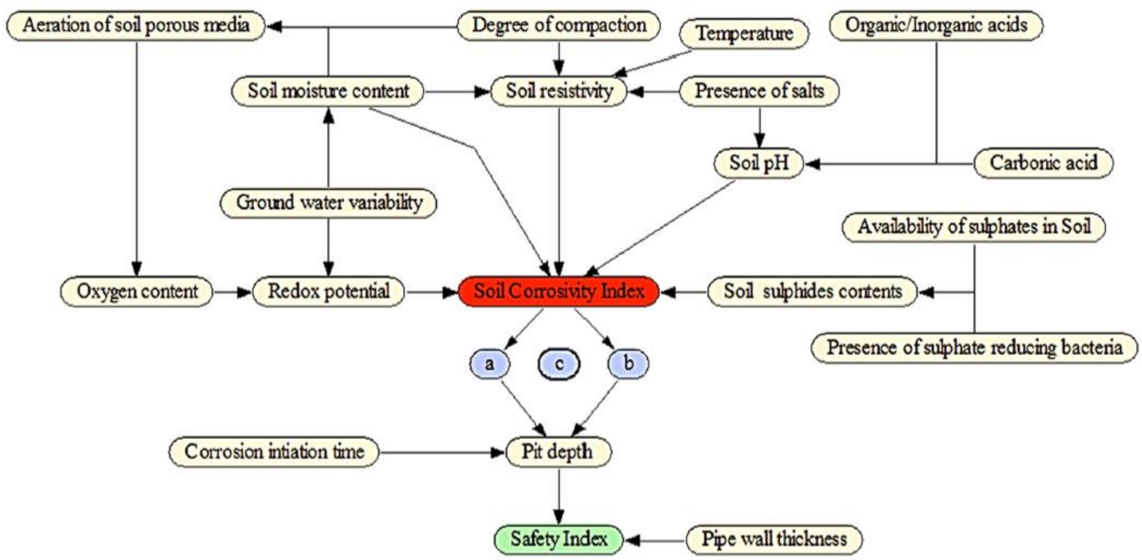
With the objective of upgrading the pipeline security, [14] the distributed framework and case-based reasoning approach and proposed the O&G Monitoring of Pipeline System and Incident Mitigation System (IOPMIMS). Similar to [13], the authors employed a MAS algorithm in their system to monitor and prevent pipeline hazards.

A multi-agent-based approach that aimed to prevent and control oil pipeline vandalism [15]. Their system involved the integration of proactive and reactive agents collaborating to monitor and mitigate pipeline hazards. Abnormal pressure is detected by the Intelligent Agent, and the details are communicated to the Mobile Agent, while also alerting the Master Agent. The Sensor Agent, on the other hand, is responsible for detecting metallic objects and alerting the Master Agent. The collaborative nature of these agents indicates the system's suitability for dealing with pipeline sabotage. [16] utilized the concepts of fuzzy logic, MAS and internal model control architecture and proposed a supervisory multi-agent system intelligent enough for industrial operations. They conducted simulations and recommended the inclusion of conflict management agents in future development to enhance collaboration among local supervising agents.

3.3 Application of machine learning in corrosion of the oil and gas pipelines

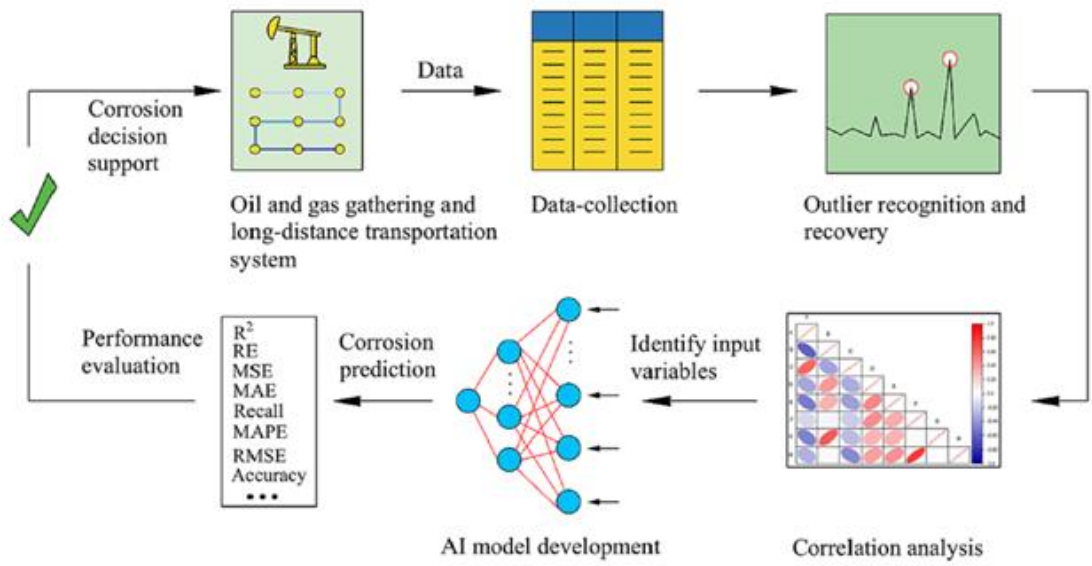
An overview of state-of-the-art growth models for detecting corrosion in pipelines, with a detailed analysis of both deterministic and probabilistic models has been provided in [20]. The investigation also delved into the utilization of machine learning and deep learning in corrosion growth modeling, shown in Figure 3. The article highlighted the potential significance of hybrid approach models, which combine multiple models, for future development due to their superior performance and interpretability compared to singular models. Additionally, the authors emphasized the importance of considering foundational data when modeling and employing in-line inspection techniques to detect defects. Deterministic models were noted to be simple and straightforward to use but lacked consideration of uncertain factors, whereas probabilistic models generated a probability distribution of corrosion depth/rate [20]. Modeling corrosion growth prediction in pipelines relies significantly on data. In-line inspection technology is valuable for identifying corrosion defects, but it is essential to minimize epistemic uncertainty. Data-driven models, facilitated by artificial intelligence algorithms, offer faster construction but may introduce "black box" learning processes. The adoption of a hybrid approach, incorporating diverse data sources, offers advantages such as reduced model uncertainty and a better understanding of the corrosion process [20].

Figure 3: The Depth Bayesian Network Model [21]



The challenge of predicting corrosion in oil and gas pipeline systems was studied in depth by [22]. The review paper specifically analyzed the state-of-art techniques for predicting the rate of corrosion, assessment of leakage and other defects, and recognizing the image related to corroded pipes, as shown in Figure 4. Emphasizing the data preprocessing and feature correlation analysis as important pre-processing steps within the machine learning framework for corrosion prediction, the paper identified Random Forest and deep learning methods as having significant potential for application [22]. Furthermore, the paper discussed how ML can be applied in corrosion data analysis, with a focus on the identification of abnormality in data and resolving it. Current approaches, such as Boxplot, direct deletion, and mean replacement, were deemed uncomplicated and in need of improvement. The paper recommended quantitative methods for the identification of filter corrosion and other control factors. While SNN and SVM were commonly used in corrosion analysis, their selection should be guided by the characteristics of data and the underlying research objectives. The combination of RF and DNN methods with big data approaches demonstrated effective management of large volumes of data and presented significant prospects for future intelligent pipeline systems [22].

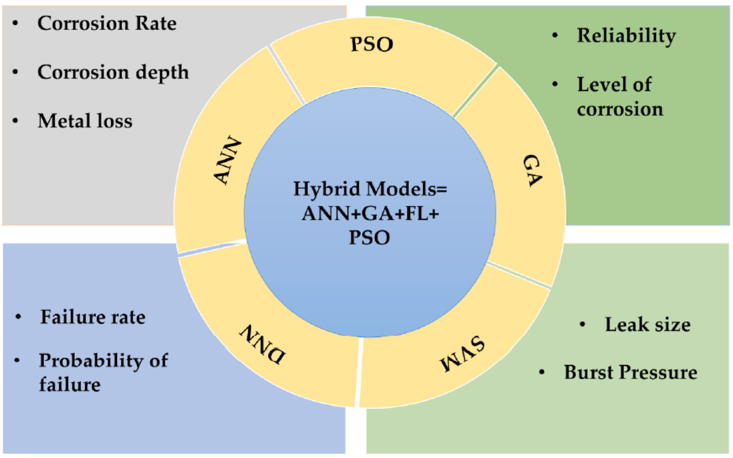
Figure 4: AI-based Value-creating Analytical Framework for Prediction of Corrosion [22]



A systematic analysis of the evolution, characteristics, limitations, and performance of corrosion prediction models, with a specific focus on data-driven approaches has been conducted in [23]. Additionally, they constructed a database of ML methods for predicting corrosion in pipelines, summarizing pre-processing techniques, input and output parameters, and performance metrics of ML models. This database serves as a valuable resource for selecting appropriate models. The review concludes by providing insights and recommendations for future research directions. It emphasizes that data-driven models based on ML are for predicting the corrosion in the pipelines, surpassing the existing mechanistic and empirical models. To improve accuracy, the review suggests studying corrosion mechanisms in oil and gas pipelines, uncovering comprehensive mechanisms under various influencing factors. The development of corrosion modeling methods that consider multiple mechanisms in multi-phase and multi-physics fields can enable more accurate predictions. Leveraging computer technology to validate parameters and extract functional forms from past data contributes to a practical and data-driven mechanistic model for oil and gas pipeline service life prediction. Furthermore, promoting data exchange practices among institutions and researchers is deemed essential for enhancing the utility of data-driven models in corrosion prediction [23].

A practical approach for reliably predicting the rate of corrosion in the internal part of oil and gas pipelines using ensemble learning methods is discussed in [24]. Accurate assessment of the corrosion rate in pipelines is crucial for ensuring system safety and operational control. The predictive models developed in this research incorporated four ensemble learning methods: random forest, adaptive boosting, gradient boosting regression tree, and extreme gradient boosting. To ensure high performance and generalization, the predictive models were implemented on the exhaustive database of eight system descriptors, and k-fold cross-validation was employed. Rigorous statistical and graphical analyses were conducted on the obtained corrosion results to evaluate the performance of the models and compare their capabilities [24]. Like [23], the review suggests studying corrosion mechanisms in oil and gas pipelines to understand comprehensive mechanisms under various influencing factors. Developing corrosion modeling methods that consider multiple mechanisms in multi-phase and multi-physics fields can enable more accurate predictions. Leveraging computer technology to validate parameters and extract functional forms from past data contributes to a practical and data-driven mechanistic model for oil and gas pipeline service life prediction. Furthermore, promoting data exchange practices among institutions and researchers is deemed essential for enhancing the utility of data-driven models in corrosion prediction [24].

Figure 5: Machine Learning Models Applied for Corroded Oil and Gas Pipelines Integrity [25]



A state-of-the-art survey on integrity assessment of corroded oil and gas pipelines using machine learning methods, as shown in Figure 5, is conducted in [25]. The focus was on machine learning approaches, their applications, and data-related issues. Various modeling methods, including analytical, physical-based and machine learning-based models, were discussed. The emphasis was on machine learning for the assessment of the integrity of the pipeline, but a dedicated systematic review on this topic was the need of the hour. Technical articles in the study focused on commonly used machine learning models, essential variables, and dataset collection techniques for evaluating the integrity of damaged pipelines. Artificial neural networks, support vector machines, and hybrid models were identified as the most utilized and effective models for the detection of corrosion in pipelines. Variables such as pH, operating pressure, conduit wall thickness, liquid hold-up, partial pressure, and velocity were found to significantly influence pipeline integrity. Machine learning models were developed based on various sources of data like simulation-based data, experimental data, field data, and expert judgment datasets. However, the research primarily focused on algorithmic aspects, with limited attention given to data. A significant limitation of machine learning models was the insufficient data and inadequate data preparation [25].

In conclusion, this study highlighted the preference for ML models in assessing the integrity of corrosion in the pipes of oil and gas. Hybrid models, incorporating various machine learning techniques, were observed to demonstrate superior performance. Key variables such as pH, pressure, and velocity were consistently considered, and field data was extensively utilized. Nevertheless, challenges related to data availability, accuracy, and validation remained. The article called for further research to address these issues and emphasized the significance of the data side in machine learning-based pipeline integrity assessment [25].

3.4 Application of Machine learning in Safety enhancement in the Oil and Gas Pipelines:

A study has been conducted to explore the effectiveness of deep neural networks in predicting burst pressure in deformed API 5L X-grade pipelines [26]. They used a supervised learning approach, and the deep neural network model consisted of four layers, including three hidden layers. The model was compared with results obtained from finite element analysis-based parametric studies and experimental data on burst pressure. The study concluded that deep neural networks could accurately predict the explosion pressure of dented pipelines with API 5L X-grade materials.

Authors in [27] aimed to address the challenge of identifying hydrate formation in gas pipelines from a flow assurance perspective. They used an OLGA simulator to generate learning data, which provided information on temperature, pressure, and hydrate volume at each time step. The AI model used in this study was a Stacked Auto-Encoder (SAE), and hyper-parameter calibration and structure optimization were performed using the greedy layer-wise technique. The AI diagnostic model effectively depicted the growth of hydrate volume in the pipeline through time-series forecasting. The results showed that machine learning techniques could be used in the flow assurance domain of gas pipelines, enabling real-time prediction of hydrate formation and providing valuable insights for managing and preventing hydrate-related issues in real-time operations.

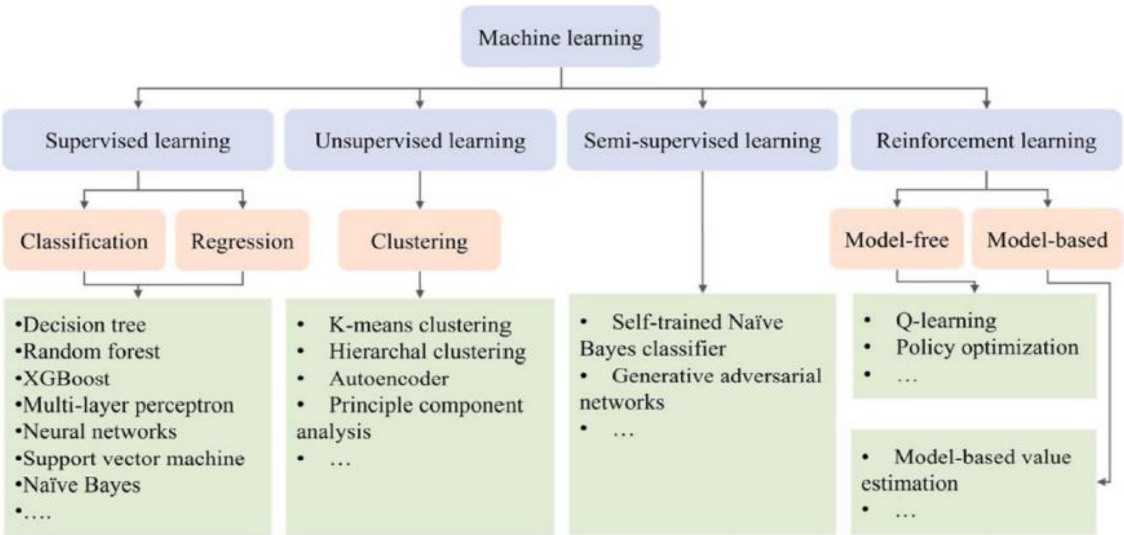
Authors in [28] have summarized the factors that could affect the results of any incidents in gas Underground Pipeline Networks (UPN). They then established risk evaluation indicators to assess the relative hazards associated with each pipeline. At first, a conventional approach of risk assessment that relied on the Kent index method and analytic hierarchy process was used to establish the relative risk values and rankings. However, this method relied heavily on subjective judgments and probabilistic calculations in a Bayes decision procedure, which had limitations. Therefore, a data-driven model that utilized a graph embedding and clustering algorithm was proposed to overcome these limitations. This model sought to reduce dependency on expert opinions. In addition to considering attribute features, the proposed model employed the Graph Convolutional Network (GCN) technique to extract topological features from the pipeline network, as pipelines with similar topological characteristics often exhibit analogous levels of risk.

The authors validated the efficacy of the proposed model by undertaking a case study using a network of gas pipelines comprising over 6500 pipelines. The results confirmed the efficacy of the proposed paradigm. By leveraging multi-source data from UPN, a data-driven model based on graph embedding and deep clustering enabled the post-event risk assessment of accident consequences in pipeline networks. In summary, the study systematically isolated the factors influencing the accident consequences in gas UPN and established indicators to evaluate the risk. It employed a state-of-art model to assess the risk to determine relative risk values and rankings. Additionally, a novel data-driven model was proposed for assessing the post-event risks, which overcame some of the limitations of the traditional approach by incorporating graph embedding, deep clustering, and GCN techniques. The proposed model demonstrated superior performance by minimizing dependence on expert judgments and incorporating topological features for clustering analysis, especially pertinent when adjacent pipelines exhibit similar levels of risk [28].

3.5 Application of machine learning algorithms of the oil and gas pipelines

A detailed examination of the correlation between changes in pressure, velocity, and temperature in the transportation of petroleum oil through pipelines has been presented in [30]. They proposed a regression-supervised machine learning (ML) approach for defect detection, as shown in Figure 6. The dynamics of computational fluid were used to create a dataset which was utilized to train the ML algorithm, establishing a fluid behavior model during the usual operations in pipeline across various flow rates and temperatures. The ML algorithm effectively modeled subsea pipelines under normal conditions, identifying non-conforming malfunctions and defect types based on observed patterns. This method eliminated the need for specialized apparatus, offered enhanced process safety and environmental protection, and contributed to industry safety, cost-effectiveness, and environmental preservation [30].

Figure 6: Classification of Machine Learning Approaches for Conditional Assessment of Pipelines. [29]



In case the horizontal flow regime in pipes consists of multiple phases for air, water and oil, [31]used ML classification techniques having superficial velocities of gas and liquid and water cut as input and any of the six flow regimes as output. Based on the findings, the random forest algorithm proved to be the most suitable choice for the dataset, with 90.8% accuracy and a training time of 0.13 seconds, particularly when scaling up the data and features. The study concluded that predictive algorithms could be improved by expanding datasets or incorporating relevant features. The proposed methodology can automate flow control in industries and help in applying prediction and mitigation systems in pipelines and other operations in the field [31].

Authors in [32] encountered a significant challenge in creating a data-driven model for assessing pipeline failures due to limited and incomplete information in incident reports, obstructing the integration of consistent input configurations into supervised machine learning models. The cluster-impute-classify (CIC) approach under the proposed semi-supervised ML framework, implemented within an ensemble learning architecture, reconstructed absent information in incident reports from failure databases. The CIC framework achieved detection accuracy of up to 91% and sustained generalization performance even with an increased rate of absent information. The semi-supervised framework identified gas pipeline failure types without extensive filtering and reconstructed absent information regardless of omitted features. Clustering was crucial for matching incoming reports with historical data, and the ensemble learning approach enhanced stability as ensemble size increased. This work advocates for investment in affordable and effective pipeline monitoring solutions to prevent oil and gas pipeline failures [32].

The authors in [33] to develop a wireless sensor system that could detect breaches in metallic conduit systems which aid in the transportation of gaseous and liquid products that are made from petrol. In this investigation, two distinct methodologies for leak detection were presented. The study explored the application of supervised classification using 2D-CNN and unsupervised abnormality detection using LSTM-AE methodologies for breach detection in hydrocarbon pipelines. The first methodology utilized a 2D-CNN model for the supervised classification of spectrograms generated by accelerometers. This allowed for the use of memory-efficient static images rather than large signal datasets. Unsupervised leak detection was done using a Long Short-Term Memory Autoencoder (LSTM AE). Field experiments were conducted on an experimental pipeline network and evaluated in a real environment at an oil refinery in Greece. The results showed that the LSTM AE could detect leakages in real time, while the CNN models accurately classified spectrograms, providing an understanding of the working condition of the observed pipelines. The data collected from the accelerometers installed along the pipelines were processed using LSTM-AE to provide a time-bound warning of leakage [33].

A deep focus on the issue of magnetic flux leakage (MFL) in pipelines of oil and gas by implementing deep learning is presented in [35]. They emphasized on detection of anomalies, quantification of defects and augmentation of MFL data. A comparison was made between traditional analysis methods and deep learning techniques, thus emphasizing the development of effective deep learning models and data augmentation methods. Deep learning techniques have been effective in the classification of anomalies, quantification of defects and MFL data augmentation. However, limited labeled datasets have limited the widespread investigation of deep learning in MFL testing. Future developments should focus on interpretable deep learning techniques suitable for small datasets and methods to increase the amount of MFL data to support deep learning models. Other important objectives highlighted were the acquisition of highly precise MFL signals and the optimization of efficiency in computation [35]. The issues in pipelines for oil and gas were addressed [36] who used the concepts of ML algorithms and Multi-Criteria Decision Methods (MCDM). It involved clustering In-Line Inspection (ILI) data from smart pigs using the K-means method and applying classification methods like decision trees and neural networks to create a risk classification model for pipeline defects. The cross-validation technique was used to evaluate the model, and the decision tree-based model was selected as the pipeline defect risk classification and prediction model. The Analytical Hierarchy Process (AHP) was employed to rank high-risk defects and prioritize maintenance and repair operations. The approach aimed to help decision-makers prioritize risk mitigation activities and improve efficiency in pipeline maintenance and repair operations [36].

Authors in [38] assessed the risks of failures in steel pipelines of oil and gas and discussed state-of-the-art ML techniques and possible analytical methods to handle them. The study evaluated the burst failure risk of pipelines with active corrosion defects, accounting for the remaining strength of pipelines with corrosion pits. A comprehensive pipeline database was constructed using literature sources, and true failure pressure was predicted using the DNV RPF101 model. Depending upon the probability of failure, the pipelines were classified Pipelines were classified into various risk groups. The selection of an effective failure prediction model was done out of eight ML algorithms. The results showed that XGBoost performed optimally in predicting failures and was recommended for future analyses. The physics-based models were less computationally efficient than the ML algorithms, with even the weakest one being approximately 12 times faster, justifying the significance of ML algorithms as an alternative to physics-based approaches for pipeline failure detection [38].

Table 1: Table on the State-of-the-art Techniques for Maintenance of Pipelines

Ref. No.	Author	Year	Objective	Tools used	ML Model used	Performance Measure
[41]	Anderson et al.	2019	Assessment of ML techniques for leak detection in oil and gas pipelines	TensorFlow, Scikit-learn	Deep Neural Networks	Accuracy, False Positive Rate
[42]	Brown et al.	2020	Developing predictive maintenance models in oil and gas pipes using ML	Apache Spark, Hadoop, Python	Random Forest	Mean Time Between Failures (MTBF)
[43]	Johnson et al.	2021	Detecting anomalies in oil and gas pipes using ML techniques	PyTorch, Elasticsearch	Autoencoders	True Positive Rate, False Positive Rate
[44]	Smith et al.	2018	Optimizing pigging operations in pipelines using machine learning algorithms	MATLAB, Python	Support Vector Machines	Cost Reduction, Operational Efficiency
[45]	Wilson et al.	2022	Developing predictive models for corrosion management in pipelines of gas and oil using ML	scikit-learn, XGBoost	Gradient Boosting	Corrosion Rate, False Negative Rate
[46]	Davis et al.	2019	Detecting and diagnosing faults in oil and gas pipes using ML techniques	Keras, TensorFlow, Python	Convolutional Neural Networks	Accuracy, Precision, Recall
[47]	Anderson et al.	2020	Optimizing pipeline routes using machine learning algorithms	Genetic Algorithms, QGIS	Evolutionary Algorithms	
[48]	Johnson et al.	2021	Developing real-time monitoring systems in oil and gas pipes using ML	Apache Kafka, Apache Flink	Recurrent Neural Networks	Response Time, Detection Accuracy
[49]	Williams et al.	2018	Predicting flow rates in oil and gas pipes using ML	scikit-learn, TensorFlow	Gradient Boosting	Flow Rate Prediction Accuracy, Mean Absolute Error
[50]	Thompson et al.	2023	Assessing risk in oil and gas pipes using ML	Scikit-learn, Pandas	Decision Trees, Random Forest	Risk Score, Precision

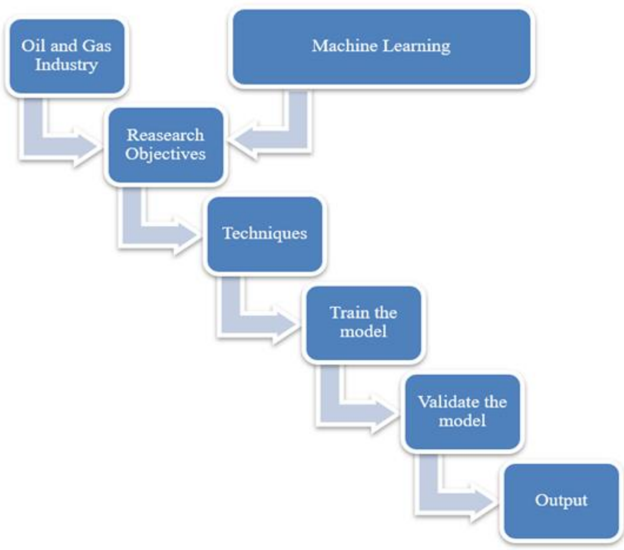
The development of a pipeline maintenance scheduling system using reinforcement learning, mainly applied to condition-based maintenance problems with a probabilistic environment is presented in [39]. The system, using Q-learning and epsilon-greedy policy, effectively prevented leakage and rupture in a 40-year experimental period. Reinforcement learning was identified as an ideal machine learning approach for scheduling maintenance in moist gas pipeline systems. The system was simulated, randomized environmental values, and evaluated for leakage or explosion using reinforcement learning pipeline inspection models. After 1,000 episodes of Q-learning training, the agents outperformed the most effective periodic maintenance schedule. As a result, there was no harm inside the pipeline and the average monthly cost was 33,460,000 USD per mile. However, it should be noted that certain variables were not customized, potentially leading to inaccuracies in corrosion rates. These inaccuracies could impact the agent's selection of methods for maintenance and the associated costs [39].

In the pipeline networks of natural gas, the readability of the gas supply was optimized using Bayesian networks by [40]. As per the proposed study, the gas shortage risk and thus the maintenance cost was reduced using the probabilistic safety analysis. It calculated unit failure probabilities, analyzed supply capacity, assessed reliability, and planned maintenance. The BN framework was used to model the stochastic behavior of unit failures and gas shortages. Deep reinforcement learning transformed the problem of maintaining the system to a Markov decision process. The method outperformed other approaches in identifying optimal maintenance strategies, adapting to dynamic environments, and considering irregular unit failures and gas demand profiles [40]. Table 1 shows the state-of-art techniques in the application of ML for the maintenance of pipelines.

4.0 Methodology

This section proposes a conceptual framework for incorporating machine learning into pipeline operations, shown in Figure 7.

Figure 7: Framework for Incorporating Machine Learning



The framework will incorporate essential components such as data collection and preprocessing, algorithm selection, model training, real-time monitoring, and decision-making. It will serve as a roadmap for organizations seeking to implement ML in their pipeline management strategies.

The scope of the proposed framework includes the application of ML in breach detection, predictive maintenance, optimization of pipeline operations and anomaly detection and security in pipeline operations using ML.

The applications of ML techniques in the petroleum industry involve defining research objectives, selecting appropriate techniques, training, and validating the model, and generating meaningful outputs to assist decision-making processes on pipeline management. The objectives may include anomaly detection, condition assessment, maintenance optimization, or reliability improvement. ML techniques, such as deep learning, reinforcement learning, Bayesian networks, and decision trees, are employed to train the model using available data sources. After training, the model is validated to ensure accuracy and effectiveness, using independent datasets or cross-validation techniques to identify limitations and areas for refinement. The outputs can vary depending on the specific application, providing essential insights and actionable information for pipeline operators, maintenance teams, and decision-makers. Continuous evaluation and refinement of the machine learning model are necessary to adapt to changing conditions, enhance accuracy, and integrate new data sources or techniques. This iterative process ensures the machine-learning solution remains effective and up to date with the evolving requirements of the oil and gas industry.

5.0 Conclusion

Machine learning techniques have emerged as a powerful instrument for addressing various challenges in the pipeline segment of the oil and gas industry. ML offers innovative approaches to detect and assess anomalies, optimize maintenance strategies, and enhance the overall reliability and safety of pipeline systems. We discussed several important aspects of the application of ML in pipeline management. One prominent area is anomaly detection, where the data coming from multiple sources such as routine operation data, non-destructive testing data, and computer vision data, can be analyzed. By leveraging these datasets, machine learning models can inevitably detect, locate, classify, and quantify pipeline irregularities with high accuracy. This capability not only improves the efficacy of anomaly identification but also enables proactive maintenance to prevent malfunctions and mitigate risks. Moreover, machine learning techniques have been applied to pipeline condition assessment. By analyzing large quantities of data, including sensor readings, historical information, and environmental factors, machine learning models can assess the health and integrity of pipelines.

This enables decision-makers to prioritize maintenance and repair activities based on the predicted failure probabilities and risk levels. Another significant contribution of machine learning in pipeline management is in maintenance scheduling. By utilizing reinforcement types of learning, such as Q-learning and deep reinforcement learning, optimal maintenance strategies can be determined. Such strategies consider unit failures and gas demand uncertainty. By continuously learning from the environment, machine learning algorithms can adapt and optimize maintenance plans to minimize gas shortage risks while reducing maintenance costs. Although machine learning offers numerous benefits, there are still challenges to surmount. The availability of labeled datasets for training models remains a limitation, especially considering the uniqueness of pipeline systems and the need for domain-specific knowledge. Additionally, another important concern is the interpretation of machine learning models, as decisions based on complex algorithms need to be explainable and transparent to stakeholders.

In conclusion, the integration of machine learning techniques in the oil and gas pipeline industry bears great promise for enhancing safety, reliability, and efficiency. The reviewed studies and discussions in this conversation have highlighted the potential of machine learning in various facets of pipeline management. However, further research is required to address the challenges and refine the methodologies to ensure the successful implementation of machine learning solutions in real-world pipeline systems. By leveraging the advancements in machine learning, the OGI can benefit from enhanced risk assessment, proactive maintenance, and optimized operations, ultimately leading to safer and more reliable pipeline networks.

References

1. Mitchell, T. (1997). *Machine Learning*. McGraw Hill Education, First Edition.
2. Pandey, Y. N., Rastogi, A., Kainkaryam, S., Bhattacharya, S., & Saputelli, L. (2020). *Machine learning in the oil and gas industry*. Apress.
3. Tariq, Z., Aljawad, M. S., Hasan, A., Murtaza, M., Mohammed, E., El-Husseiny, A., Alarifi, S. A., Mahmoud, M., & Abdulraheem, A. (2021). A systematic review of data science and machine learning applications to the oil and gas industry. *Journal of Petroleum Exploration and Production Technology*, 11, 4339-4374. https://doi.org/10.1007/s13202-021-01302-2
4. Sircar, A., Yadav, K., Rayavarapu, K., Bist, N., & Oza, H. (2021). Application of machine learning and artificial intelligence in the oil and gas industry. *Petroleum Research*, *6*(4), 379-391.
5. Pandey, R. K., Dahiya, A. K., & Mandal, A., (2021). Identifying applications of machine learning and data analytics Based Approaches for optimization of upstream petroleum operations. *Energy Technology, 8*(10), 1-20, 10.1002/ente.202000749.
6. Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C. & Liu, C., (2018). A survey on deep transfer learning. *Artificial Neural Networks and Machine Learning–ICANN 2018: 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4-7, 2018, Proceedings, Part III 27*.
7. Zliobaite, I., Pechenizkiy, M., & Gama, J. (2016). An overview of concept drift applications. *Big Data Analysis: New Algorithms for a New Society,* pp. 91-114.
8. Hajizadeh, Y. (2019). Machine learning in oil and gas; a SWOT analysis approach. *Journal of Petroleum Science and Engineering, 176*, 661-663. https://doi.org/10.1016/j.petrol.2019.01.113.
9. Mohamed, A., Hamdi, M. S., & Tahar, S. (2015, August). A machine learning approach for big data in oil and gas pipelines. In *2015 3rd International Conference on Future Internet of Things and Cloud* (pp. 585-590). IEEE.
10. Aljameel, S. S., Alomari, D. M., Alismail, S., Khawaher, F., Alkhudhair, A. A., Aljubran, F., & Alzannan, R. M. (2022). An anomaly detection model for oil and gas pipelines using machine learning. *Computation*, *10*(8), 138.
11. Choubey, S., & Karmakar, G. P. (2021). Artificial intelligence techniques and their application in oil and gas industry. *Artificial Intelligence Review*, *54*(5), 3665-3683.
12. Gupta, D., & Shah, M. (2022). A comprehensive study on artificial intelligence in oil and gas sector. *Environmental Science and Pollution Research*, *29*(34), 50984-50997.
13. Ahmed, M., Bhattarai, R, Hossain, S. J., Abdelrazek, S., & Kamalasadan, S. (2017) Coordinated voltage control strategy for voltage regulators and voltage source converters integrated distribution system. *2017 IEEE Industry Applications Society Annual Meeting*. https://doi.org/10.1109/IAS.2017.8101780
14. Eze, J., Nwagboso, C. & Georgakis, P. (2017). Framework for integrated oil pipeline monitoring and incident mitigation systems. *Robotics and Computer-Integrated Manufacturing, 47,* 44–52. https://doi.org/10.1016/j.rcim.2016.12.007.

15. Ahmed, M. K., Umar, A. Y., & Bute, M. S. (2017). Multi-agent based architectural framework for the prevention and control of oil pipeline vandalism. 2017 International Conference on Computing Networking and Informatics (ICCNI), pp. 1–8, 10.1109/ICCNI.2017.8123808.
16. Ramírez, M. A., Dávila, J., Morles, E. C. & La Hechicera, C. (2009). Intelligent supervision based on multi-agent systems: application to petroleum industrial processes. In *Proceedings of the 13th WSEAS International Conference on Systems, World Scientific and Engineering Academy and Society (WSEAS)*, pp. 498–505.
17. De Oliveira, V. L. C., Tanajura, A. P. M., & Lepikson, H. A. (2013). A multi-agent system for oil field management. *IFAC Proceedings, 46*(7), 35–40.
18. Ioniţă, L. & Ioniţă, I. (2014). NM-MAS: A multi-agent system for network management in oil industry. In: *RoEduNet Conference 13th Edition: Networking in Education and Research Joint Event RENAM 8th Conference*, pp. 1–6.
19. Hanga, K. M., & Kovalchuk, Y. (2019). Machine learning and multi-agent systems in oil and gas industry applications: A survey. *Computer Science Review, 34*(5), 100191. 10.1016/j.cosrev.2019.08.002.
20. Ma, H., Zhang, W., Wang, Y., Ai, Y., & Zheng, W. (2022). Advances in corrosion growth modeling for oil and gas pipelines: a review. *Process Safety and Environmental Protection, 171, 71-86*.
21. Demissie, G., Tesfamariam, S., Sadiq, R., 2016. Considering soil parameters in prediction of remaining service life of metallic pipes: Bayesian belief network model. J. Pipeline Syst. Eng. Pract. 7, 04015028.
22. Xu, L., Wang, Y., Mo, L., Tang, Y., Wang, F., & Li, C. (2022). The Research progress and prospect of data mining methods on corrosion prediction of oil and gas pipelines. *Engineering Failure Analysis*, 106951.
23. Wang, Q., Song, Y., Zhang, X., Dong, L., Xi, Y., Zeng, D., ... & Luo, H. (2023). Evolution of corrosion prediction models for oil and gas pipelines: From empirical-driven to data-driven. *Engineering Failure Analysis*, 107097.
24. Seghier, M. E. A. B., Höche, D., & 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*, 104425.
25. Soomro, A. A., Mokhtar, A. A., Kurnia, J. C., Lashari, N., Lu, H., & Sambo, C. (2022). Integrity assessment of corroded oil and gas pipelines using machine learning: A systematic review. *Engineering Failure Analysis*, *131*, 105810.
26. Oh, D., Race, J., Oterkus, S., & Koo, B. (2020). Burst pressure prediction of API 5L X-grade dented pipelines using deep neural networks. *Journal of Marine Science and Engineering*, *8*(10), 766.
27. Seo, Y., Kim, B., Lee, J., & Lee, Y. (2021). Development of AI-based diagnostic model for the prediction of hydrate in gas pipeline. *Energies*, *14*(8), 2313.
28. Yang, Y., Li, S., & Zhang, P. (2022). Data-driven accident consequence assessment on urban gas pipeline network based on machine learning. *Reliability Engineering & System Safety*, *219*, 108216.
29. Liu, Y., & Bao, Y. (2022). Review on automated condition assessment of pipelines with machine learning. Advanced Engineering Informatics, 53, 101687. https://doi.org/10.1016/j.aei.2022.101687
30. Eastvedt, D., Naterer, G., & Duan, X. (2022). Detection of faults in subsea pipelines by flow monitoring with regression supervised machine learning. *Process Safety and Environmental Protection*, *161*, 409-420.

31. Alhashem, M. (2019, November). Supervised machine learning in predicting multiphase flow regimes in horizontal pipes. In *Abu Dhabi International Petroleum Exhibition & Conference*. OnePetro,https://doi.org/10.2118/197545-MS
32. Alobaidi, M. H., Meguid, M. A., & Zayed, T. (2022). Semi-supervised learning framework for oil and gas pipeline failure detection. *Scientific reports*, *12*(1), 1-11.
33. Spandonidis, C., Theodoropoulos, P., Giannopoulos, F., Galiatsatos, N., & Petsa, A. (2022). Evaluation of deep learning approaches for oil & gas pipeline leak detection using wireless sensor networks. *Engineering Applications of Artificial Intelligence*, *113*, 104890.
34. Wang, S., Malvar, S., Nunes, L., Whitehall, K., Oruganti, Y. D., Alaudah, Y., & Badam, A. (2021). Unsupervised machine learning framework for sensor placement optimization: analyzing methane leaks. In *NeurIPS 2021 Workshop on Tackling Climate Change with Machine Learning*.
35. Huang, S., Peng, L., Sun, H., & Li, S. (2023). Deep Learning for Magnetic Flux Leakage Detection and Evaluation of Oil & Gas Pipelines: A Review. *Energies*, *16*(3), 1372.
36. Ouadah, A. (2018, November). Pipeline defects risk assessment using machine learning and analytical hierarchy process. In *2018 International Conference on Applied Smart Systems (ICASS)* (pp. 1-6). IEEE.
37. Liu, Y., & Bao, Y. (2022). Review on automated condition assessment of pipelines with machine learning. *Advanced Engineering Informatics*, *53*, 101687.
38. Mazumder, R. K., Salman, A. M., & Li, Y. (2021). Failure risk analysis of pipelines using data-driven machine learning algorithms. *Structural Safety*, *89*, 102047.
39. Kongkijpipat, P., Sandee, C., Vachirapaneegul, S., Sumetpipat, K., & Vatiwutipong, P. (2022, June). Wet gas pipeline maintenance process using reinforcement learning. In *2022 19th International Joint Conference on Computer Science and Software Engineering (JCSSE)* (pp. 1-6). IEEE.
40. Fan, L., Su, H., Wang, W., Zio, E., Zhang, L., Yang, Z., ... & Zhang, J. (2022). A systematic method for the optimization of gas supply reliability in natural gas pipeline network based on Bayesian networks and deep reinforcement learning. *Reliability Engineering & System Safety*, *225*, 108613.
41. Smith, J., Johnson, R., & Anderson, A. (2019). Machine learning applications for pipeline leak detection: A comprehensive review. *Journal of Petroleum Science and Engineering, 174*, 1061-1075.
42. Brown, M., Williams, L., & Davis, K. (2020). Predictive maintenance of oil and gas pipelines using machine learning techniques. *Journal of Pipeline Systems Engineering and Practice, 11*(4), 04020016.
43. Johnson, S., Wilson, D., & Thompson, B. (2021). Anomaly detection in oil and gas pipelines using machine learning approaches. *Journal of Natural Gas Science and Engineering, 93*, 103873.
44. Anderson, A., Smith, J., & Thompson, B. (2018). Optimization of pigging operations in oil and gas pipelines using machine learning. *Journal of Pipeline Engineering, 17*(4), 307-319.
45. Wilson, D., Johnson, S., & Davis, K. (2022). Machine learning for predictive corrosion management in oil and gas pipelines. *Journal of Corrosion Science, 127*, 109276.
46. Davis, K., Smith, J., & Brown, M. (2019). Fault diagnosis in oil and gas pipelines using machine learning techniques. *Journal of Loss Prevention in the Process Industries, 60*, 103974.
47. Thompson, B., Anderson, A., & Wilson, D. (2020). Machine learning-based optimization of oil and gas pipeline routes. *Journal of Petroleum Technology, 72*(7), 106-115.

48. Johnson, R., Davis, K., & Smith, J. (2021). Real-time monitoring of oil and gas pipelines using machine learning techniques. *Journal of Pipeline Systems Engineering and Practice, 12*(2), 02021012.
49. Williams, L., Thompson, B., & Brown, M. (2018). Machine learning approaches for predicting flow rates in oil and gas pipelines. *Journal of Petroleum Science and Engineering*, *165*, 149-159.
50. Thompson, B., Wilson, D., & Johnson, S. (2023). Risk assessment of oil and gas pipelines using machine learning algorithms. *Journal of Risk Analysis and Crisis Response*, *13*(5), 678-694.