Prediction of oil and gas pipeline failures through machine learning approaches

A systematic review

Al-Sabaeei, Abdulnaser M.; Alhussian, Hitham; Abdulkadir, Said Jadid; Jagadeesh, Ajayshankar

DOI
10.1016/j.egyr.2023.08.009

Publication date
2023

Document Version
Final published version

Published in
Energy Reports

Citation (APA)
Al-Sabaeei, A. M., Alhussian, H., Abdulkadir, S. J., & Jagadeesh, A. (2023). Prediction of oil and gas pipeline failures through machine learning approaches: A systematic review. *Energy Reports*, *10*, 1313-1338. https://doi.org/10.1016/j.egyr.2023.08.009

Energy Reports 10 (2023) 1313–1338

Contents lists available at ScienceDirect

Energy Reports

journal homepage: www.elsevier.com/locate/egyr

Review Article

Prediction of oil and gas pipeline failures through machine learning approaches: A systematic review

Abdulnaser M. Al-Sabaeei [a,b,*], Hitham Alhussian [b], Said Jadid Abdulkadir [b], Ajayshankar Jagadeesh [c]

[a] Department of Civil & Environmental Engineering, Universiti Teknologi PETRONAS, 32610 Bandar Seri Iskandar, Perak, Malaysia
[b] Centre for Research in Data Science (CeRDaS), Computer Information Science Department, Universiti Teknologi PETRONAS, 32610 Seri Iskandar, Perak, Malaysia
[c] Faculty of Civil Engineering and Geosciences, Delft University of Technology, 2628 CD Delft, Netherlands

A R T I C L E I N F O

Article history:
Received 12 March 2023
Received in revised form 30 July 2023
Accepted 2 August 2023
Available online xxxx

Keywords:
Energy transportation system (pipeline)
Oil and gas
AI algorithms (machine learning)
Advanced neural networks

A B S T R A C T

Pipelines are vital for transporting oil and gas, but leaks can have serious consequences such as fires, injuries, pollution, and property damage. Therefore, preserving pipeline integrity is crucial for a safe and sustainable energy supply. The rapid progress of machine learning (ML) technologies provides an advantageous opportunity to develop predictive models that can effectively tackle these challenges. This review article mainly focuses on the novelty of using machine and deep learning techniques, specifically artificial neural networks (ANNs), support vector machines (SVMs) and hybrid machine learning (HML) algorithms, for predicting different pipeline failures in the oil and gas industry. In contrast to existing noncomprehensive reviews on pipeline defects, this article explicitly addresses the application of ML techniques, parameters, and data reliability for this purpose. The article surveys research in this specific area, offering a coherent discussion and identifying the motivations and challenges associated with using ML for predicting different types of defects in pipelines. This review also includes a bibliometric analysis of the literature, highlighting common ML techniques, investigated failures, and experimental tests. It also provides in-depth details, summarized in tables, on different failure types, commonly used ML algorithms, and data resources, with critical discussions. Based on a comprehensive review aforementioned, it was found that ML approaches, specifically ANNs and SVMs, can accurately predict oil and gas pipeline failures compared to conventional methods. However, it is highly recommended to combine multiple ML algorithms to enhance accuracy and prediction time further. Comparing ML predictive models based on field, experimental, and simulation data for various pipeline failures can establish reliable and cost-effective monitoring systems for the entire pipeline network. This systematic review is expected to aid in understanding the existing research gaps and provide options for other researchers interested in predicting oil and gas pipeline failures.

Contents

* Corresponding author at: Department of Civil & Environmental Engineering, Universiti Teknologi PETRONAS, 32610 Bandar Seri Iskandar, Perak, Malaysia.
E-mail address: abdulnaser_17005477@utp.edu.my (A.M. Al-Sabaeei).

https://doi.org/10.1016/j.egyr.2023.08.009
2352-4847/

Abbreviations

1DCNN	One-dimensional neural network
AI	Artificial intelligence
ANNs	Artificial neural networks
CCTV	Closed circuit TV
CLR	Conventional literature review
CNN	Convolutional neural network
CO_2	Carbon dioxide
CPSO	Chaos particle swarm optimization
DNN	Deep neural network
EAC	Environmental aid cracking
FEM	Finite element method
FFNN	Feed-forward neural network
GP	Genetic programming
HML	Hybrid machine learning
ICA	Independent component analysis
ILI	Inline inspection
KNN	K-nearest neighbors
LR	Liner regression
LS-SVM	Least-square support vector machine
LSTM-AE	Long short-term memory autoencoder
MFL	Magnetic flux leakage
MIC	Microbiologically influenced corrosion
ML	Machine learning
MLP	Multiple layer perceptron
NNs	Neural networks
OCSVM	One-class support vector machine
OLGA	Oil & gas
PCA	Principle component analysis
pH	Potential hydrogen
PIM	Pipeline integrity management
PLS	Partial least squares
POF	Probability of failure
PPA	Posterior probability of association
PRISMA	Preferred reporting items for systematic reviews and meta-analyses
PSEW	Pipeline safety early warning
RBFN	Radial basis function network
RFR	Random forest regression
RRSE	Root relative square error
SAM	Similarity aggregation method
SCC	Stress corrosion cracking
SCNNs	Subspace clustered neural networks
SLR	Systematic literature review
SPPS	Sand production pipe saver
SPRT	Sequential probability ratio test
SSA-CNN	Sparrow Search Algorithm and Convolutional Neural Network
SVMs	Support vector machines
TOC	Top of the line corrosion
UK	United Kingdom
USA	United States of America
USD	United States dollar
VAPSO	Variable amplitude particle swarm optimization
VBA	Visual basic for applications
VOSviewer	Visualization of similarities viewer
XGBoost	Extreme Gradient Boosting

1. Introduction

The growing global oil and gas industry since the middle of the nineteenth century led to high demand for transportation and linking oil and gas to markets. Pipelines made of stainless steel are considered the most economical and advanced technology that is currently utilized for oil and gas transportation and there are about 2.2 million miles of pipelines around the world (Seghier et al., 2022; Valentin de Oliveira, 2018; Vandrangi et al., 2022). The pipelines play essential roles in oil and gas transportation, especially with the rapid development of societies and increasing demand for oil and gas (Liao et al., 2022; Zhou et al., 2022). Compared to conventional transportation methods, pipelines are considered fast, low-cost, easier for operating and transport a large volume of oil and gas. Thus, the majority of the oil and gas industry worldwide uses pipelines as the main transportation way for oil and gas (Liu et al., 2020, 2019a; Peng et al., 2021; Zaman et al., 2020). Pipelines are also the safest means of oil, gas and refined petroleum products transportation (Liu et al., 2023, 2022; Shaik et al., 2022). Millions of kilometers of pipelines have been constructed worldwide to fulfill the continuous demand for oil gas products (Agency, 2020; Khan et al., 2021). Despite the various advantages of utilizing pipelines compared to other oil and gas transportation methods in terms of low cost, fast and easier; but the failures such as leakage associated with the use of pipelines result in notable challenges for the oil and gas industries worldwide (Adegboye et al., 2019; Vandrangi et al., 2022; Wu et al., 2023; Zhang et al., 2022).

Oil and gas pipelines are likely to leak due to various parameters such as operating conditions (including the aggressive medium and overpressure), surrounding environment (including atmosphere, soil, earthquake and flood) and human factors (Such as excavation, bad installation and oil stolen) (Yin et al., 2021). One example of such pipeline accidents that occur in various countries is that about 745 major accidents only in the USA between 1994 and 2013 caused 278 deaths, 1059 injuries, and 110 million USD as a monetary loss (Vandrangi et al., 2022). In order to mitigate the risks of the defects and failures that can be introduced along and across the oil and gas pipelines and to avoid the risk toward the societies, environment and oil and gas industries workers, regular assessments and predictions of the failures prior to occurring should be conducted (Biezma et al., 2020; Khan et al., 2021; Zheng et al., 2022; Zuo et al., 2022). Besides, the quick detection the oil and gas pipeline failures or leakage is crucial to minimize maintenance and repair expenses (Li et al., 2022a; Yao et al., 2022). Early detection and maintenance of oil and gas pipeline failures are also necessary to avoid needless loss faced by oil and gas industries and also to maintain a safe environment for different work conditions.

Although the advantages of using pipelines to transport oil and gas compared to other methods such as ship, road and rail transportation, it is still facing serious failures issues that can lead to severe negative consequences on equipment and properties of transmission medium such as explosion, toxicity and flammability and the effect on the environment due to the complex service conditions such as internal corrosion, atmosphere corrosion, external soil and third-party damage. Therefore, the accurate evaluation of the failures criticality of oil and gas pipelines is one of the most important parameters that can guarantee the long-term safe and economic service of pipeline networks (Crawley, 2020; Fang et al., 2020; Girgin and Krausmann, 2016; Yin et al., 2021). In general, the criticality analysis of oil and gas pipelines can be conducted using three different methods: quantitative, qualitative and semi-quantitative. The quantitative method is mainly conducted based on materials properties, process situations and physical models that describe the development of accident scenarios. Numerical simulation is one of the common examples of quantitative analysis used. On the other hand, qualitative analysis mainly depends on the estimations given by engineers or managers which reflect the people's experience and intuitions play the main role in such estimations (Yin et al., 2021).

The quantitative method may come up with better assessment results compared to the qualitative method, but the quantitative technique requires precise and complete data, more cost and more time. Unfortunately, most of the time it is quite challenging to get the field data that reflect the reliable failure records due to the complex operating conditions. Therefore, the semi-quantitative method was suggested which can overcome the limitations of the two techniques because of its high flexibility and adequate for being used in the absence of complete data and an accurate physical model. In order to carry out a semi-quantitative analysis, several attempts have been conducted using Fuzzy set theory which is a very common method to deal with incompleteness and uncertainty. However, the fuzzy method is considered relatively time-consuming and required complex inference and operations processes. Alternatively, the quick evolution of machine learning (ML) technologies provides a potential way to address the aforementioned problems by learning the mapping from data (Hegde and Rokseth, 2020; Worrell et al., 2019). It was also stated that due to the effects of internal complex medium and external aggressive environment, Oil and gas pipelines are susceptible to failure. Direct quantitative evaluation of oil and gas pipeline failures is considered very difficult because of the uncertainty and complexity of the failure scenarios of such pipelines. Therefore, advanced techniques such as machine learning are recommended to evaluate oil and gas pipeline failures (Spandonidis et al., 2022; Yin et al., 2021).

Machine learning (ML) algorithms are widely recognized as the predominant approach for developing predictive models in complex engineering, energy and environmental problems (Amini et al., 2023; Chen et al., 2022; Du et al., 2023; Vadyala et al., 2022). ML has the power to enhance the quality of predictivity, reduce the dependence on conventional and manual data analysis, provide autonomous information processing and assist in evaluating and managing high variety, velocity and volume data (Loebbecke and Picot, 2015; Rachman et al., 2021). The increasing popularity of ML in diverse fields can also be attributed to its remarkable ability to learn and construct predictive models based on performance, even when working with incomplete and empirical data (Chen et al., 2022; Li et al., 2022b; Murphy, 2012). ML algorithms possess the capability to effectively address complex nonlinear problems (Chen et al., 2022; Ma et al., 2023; Murphy, 2012). ML approaches provide a big advantage by being able to understand complex patterns without needing any prior knowledge of how the independent and dependent variables are related (Behnood and Daneshvar, 2020; Sun et al., 2022). ML algorithms are suitable for predicting the performance of engineering materials, like oil and gas pipelines, because they can accurately and swiftly estimate mechanical properties at a lower cost compared to traditional modeling methods (Du et al., 2023). ML algorithms used for detecting defects in pipelines can be classified into different categories, including supervised, semi-supervised, unsupervised, or reinforcement learning, depending on the learning method employed (Alamri, 2022; Liu and Bao, 2022; Rachman et al., 2021). In recent studies focused on detecting defects in oil and gas pipelines using ML applications, the supervised learning technique has emerged as the most widely utilized approach (Eastvedt et al., 2022; Liu and Bao, 2022; Rachman et al., 2021). Supervised machine learning involves using different algorithms to analyze datasets and identify patterns and predictions for future values (Murphy, 2012). The supervised machine learning models developed for detecting defects in oil and gas pipelines are classified into two categories: classification models and regression models. The choice between these categories depends on the primary objective of the model, which includes identifying the types of defects and predicting various aspects such as dimensions, pressure values, severity, and more (Liu and Bao, 2022).

Due to the complicated nature and vast scale of oil and gas pipeline systems, relying on human operators to carry out inspections proves to be a difficult and expensive approach. For many years, oil and gas industries, governments and researchers are looking for automating the inspection process to enhance the quality of inspection and reduce the efforts, cost and environmental consequences due to relying on humans. However, several techniques have been investigated to overcome this problem such as analytical modeling, numerical computations and machine learning (Layouni et al., 2014). Machine learning become a hot research area have been applied in various fields of life and oil and gas pipeline inspection is one of these fields. However, most research still depends on probabilistic models to predict the failures behavior of oil and gas pipelines but ML-based approaches are considered one of the best choices for anticipating unexpected failures of oil and gas pipelines, particularly due to the complexity and extensive nature of oil and gas pipelines failures (Soomro et al., 2022a). The utilization of machine learning techniques was employed to automatically detect leakage defects by analyzing pressure data and accurately determining the location of the leakage through the analysis of sound data (da Cruz et al., 2020; Zhou et al., 2021). Furthermore, computer vision methods were employed to evaluate pipeline defects by incorporating images, numeric data, and videos as additional sources of information (Wang and Cheng, 2020). In general, recent research has demonstrated the promising potential of machine learning techniques in automatically detecting, localizing, and classifying pipeline defects (Liu and Bao, 2022). It was also stated that using ML to assess the safety of oil and gas pipelines is a new research field and there is a lack of literature on the comprehensive assessment of current research issues (Soomro et al., 2022a).

On the other hand, it was claimed that traditional machine learning models that solely rely on available data without incorporating engineering theory will have limitations in terms of accuracy and efficiency (Wang et al., 2020). Model selection and parameter optimization are also crucial components in the development of a prediction model (Ramadhan et al., 2021). Most previous studies have manually chosen models and optimized their parameters, which has made it difficult to determine the best model structure and has also increased the time needed to prepare for predicting defects (Du et al., 2023). To address these limitations, it was stated that the integration of engineering theories and domain knowledge into machine ML models is emerging as a promising approach for solving various engineering problems (Du et al., 2023). For instance, in their work, Karpatne et al. (2017) introduced a method called theory-guided data science (TGDS), which combines scientific principles with data science techniques. In this regards, Du et al. (2023) conducted a recent study to introduce an automatic machine learning (AML) for optimizing the development of a corrosion depth prediction model. To address the limitations of traditional ML modeling methods, the integration of engineering theory and domain knowledge was incorporated into the feature engineering stage. Furthermore, a new prediction method, referred to as theory-guided AML (Tg-AML), was proposed specifically for accurately predicting the maximum depth of pitting corrosion in pipelines. Based on the aforementioned, researchers have conflicting opinions on the reliability and accuracy of using ML for pipeline defects detection. While some argue that ML is a dependable method (Bastian et al., 2019; Liu et al., 2021; Santoso et al., 2014), others emphasize the importance of considering various factors, such as data pre-processing, cleaning, automated algorithm selection, feature extraction and combinations and integrating with other most engineering common techniques, to enhance the accuracy and validation of ML models (Lu et al., 2021; Qin et al., 2023; Xiao et al., 2019; Xu et al., 2023; Zhou et al., 2021). Therefore, to help researchers interested in using ML techniques for detecting defects in oil and gas pipelines, a thorough and organized review is necessary. This will simplify the understanding of different viewpoints from various researchers in the literature.

Numerous advantages can be derived from developing systematic reviews in certain research areas. For instance, the review of reliable published works will aid other researchers to have a clear image of what needs to be done in this research area. It also helps researchers to compare studies on a certain topic from different perspectives to come up with meaningful arrangements. The structure of a literature review also can be used to present a clear insight in several ways, as well as direct and manage interested researchers. This systematic literature review (SLR) distinguished itself in the selection process of evaluated studies through systematic, direct research and transparency to minimize the common errors in the data collection for conventional reviews. Despite the several review articles on oil and gas pipeline defects, only very few mention the use of machine and deep learning techniques in the early prediction of oil and gas pipeline failures. However, based on the authors' observation, there has so far been no comprehensive and coherent review in order to highlight the use of machine learning techniques, specifically artificial neural networks (ANNs), support vector machines (SVMs) and hybrid machine learning (HML) algorithms, parameters used, reliability of the data in the prediction of different oil and gas pipeline failures. Therefore, the purpose of this systematic review paper is to address the lack of clarity among researchers regarding the selection of machine learning (ML) models, parameters, and reliable data resources for predicting oil and gas pipeline defects. It provides a comprehensive overview of research achievements in using ML methods to predict pipeline failures in the oil and gas industry. The paper compares and analyzes the advantages of common ML methods in detecting pipeline failures. Its value lies in assisting other researchers interested in this field by providing insights into existing research gaps and aiding in the informed selection of ML models, parameters, and data resources.

This review article specifically focuses on the use of ANNs, SVMs, and HML algorithms for predicting various pipeline failures in the oil and gas industry. It goes beyond a mere survey of research by offering a coherent discussion that identifies motivations, challenges, and future directions associated with using ML for predicting different types of defects in oil and gas pipelines. The paper explores the reasons behind employing ML techniques for predicting pipeline failures, such as early detection and improving safety and reliability. It also suggests ways to address implementation challenges, including data quality, algorithm selection, and parameter optimization. Moreover, the article offers insights into future directions, such as integrating multiple ML algorithms, using real-time data for continuous monitoring, and employing advanced techniques like deep learning for complex defect detection. Additionally, the paper includes a bibliometric analysis of the literature, highlighting common ML techniques used in pipeline failure prediction, types of failures investigated, and experimental tests conducted. This analysis helps identify trends and patterns in the research landscape, providing valuable information for researchers and practitioners. The paper also includes tables summarizing different failure types, commonly used ML algorithms, and available data resources, along with critical discussions. These discussions further enhance the value of this systematic review paper.

2. Methodology

The systematic literature review (SLR) is a potential method of literature that is commonly used to discover, highlight, assess and analyze the efforts of researchers on a particular research area of interest (Soomro et al., 2022a). Many benefits can be derived from conducting SLR such as reliable published works in certain research areas can be summarized. For example, new researchers in the applications of ML for oil and gas pipeline failure can feel confused to select the most appropriate ML approach for certain pipeline defects that can result with adequate accuracy, especially with the number of papers published on this topic without any structures, organization and analysis. Besides, which data collection method could be more useful and which parameters have the influence on the certain pipeline defect that need to be predicted. Therefore, the following SLR method and procedures have been used to come up with a reliable, structured and organized summary of the applications of ML for pipeline defects prediction. Table 1 presents a comparison of the advantages and disadvantages of systematic and conventional methods for conducting literature reviews. It highlights how the systematic literature review (SLR) is considered the most suitable option for this review based on its confirmed benefits. Overall, systematic reviews provide a more rigorous, transparent, and comprehensive method of reviewing the literature, resulting in more reliable and trustworthy conclusions compared to conventional reviews.

Based on the advantages and disadvantages of the most common methods presented in Table 1 for conducting a literature review, it can be concluded that systematic reviews are generally considered superior to conventional reviews. This is due to their thoroughness and lack of bias, resulting in reliable results. In contrast, conventional reviews may cover more sources but are often biased and less objective. Therefore, the systematic literature review method has been adopted for this study.

2.1. Databases and keywords

To find relevant research articles on the applications of ML for oil and gas pipeline failure predictions, three databases have

Table 1
Advantages and disadvantages of systematic and conventional literature review methods summarized.

Method	Advantages	Disadvantages
Systematic literature review (SLR)	SLR uses a structured methodology to ensure unbiased and transparent identification, selection, and analysis of relevant studies (Briner and Denyer, 2012; Wright et al., 2007).	Time-consuming due to the need for thorough searching, screening, and analysis of numerous studies (Wright et al., 2007).
	SLR ensures comprehensive and unbiased coverage of relevant studies by conducting exhaustive searches, minimizing the risk of overlooking important studies (Rother, 2007; Wright et al., 2007).	It demands significant resources, including personnel, funding, and access to databases and research materials (Wright et al., 2007).
	SLR employs criteria to select studies and extract data, promoting objectivity and enabling replication. This allows researchers to assess thoroughness and replicate if needed (Briner and Denyer, 2012; Wright et al., 2007)	SLR can be biased towards positive or significant findings due to their reliance on published studies, potentially overlooking unpublished results (Wright et al., 2007; Xiao and Watson, 2019).
	SLR identifies research gaps, guiding future studies and identifying areas needing further investigation (Wright et al., 2007).	SLR requires expertise in research methodology, statistics, and data analysis, making it challenging for non-specialized researchers to conduct a precise review (Wright et al., 2007).
	SLR provides a reliable summary of the evidence, helping policymakers and stakeholders make informed decisions and develop evidence-based guidelines (Rother, 2007; Xiao and Watson, 2019).	
Conventional literature review (CLR)	CLR reviews provide flexibility in scope and inclusion criteria, allowing researchers to include relevant studies or sources that align with their research question without strict criteria (Rother, 2007).	Lack of a systematic and rigorous approach to searching for and selecting relevant studies (Xiao and Watson, 2019).
	CLR usually takes less time and requires fewer resources than a SLR (Wright et al., 2007).	Higher likelihood of excluding crucial evidence due to the lack of predefined inclusion criteria (Xiao and Watson, 2019).
	CLR enables the exploration of diverse sources, such as non-peer-reviewed articles, books, opinions, and grey literature (Rother, 2007).	Researchers are more likely to be biased and subjective when they selectively include studies that confirm their pre-existing beliefs (Wright et al., 2007; Xiao and Watson, 2019).
		The lack of a clear and replicable methodology makes it difficult for others to reproduce or verify the findings (Rother, 2007; Xiao and Watson, 2019).
		The lack of explicit methods to minimize bias can lead to skewed or unreliable results (Rother, 2007).
		The lack of transparency and reproducibility in the study process can affect the reliability of the results (Wright et al., 2007).

been used: Web of Science, Scopus and Science Direct which are found to have the majority of materials on this research area based on the preliminary survey. After surveying the abstracts of the majority of articles on this research subject and discuss with experts, the keywords were selected based on the most frequently used and based on the aim of this SLR. The search queries were developed using the following words: "Pipeline" AND "Oil and gas" AND ("machine learning" OR "deep learning" OR "neural network" OR "support vector") AND ("failure" OR "defect" OR "corrosion" OR "leak" OR "collision"). The search included articles from journals and conference proceedings that were published in the English language within the span of time from January 2000 to December 2022.

2.2. Inclusion and exclusion criteria

Overall, this SLR was carried out with respecting the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) criteria (Moher et al., 2009) paired with a bibliometric analysis approach. Besides, there are four inclusion criteria that have been implemented in this review which were developed in response to the aim of this study. That includes the articles that reviewed the applications of ML in oil and gas pipelines failures prediction, the articles that developed ML-based models for oil and gas pipeline defects prediction and assessment, the articles that are peer-reviewed and the articles that are in English. Articles that have not directly met the aforementioned criteria were excluded.

2.3. Data acquisition and filtration

The quarry search has been aforementioned resulting in 255 articles: 45 from Web of Science, 72 from ScienceDirect and 138 from Scopus. After screening out the duplicated studies, the scanning of the titles and abstracts of the articles was conducted. A total of 65 articles that met the inclusion criteria were included after the full-text reading. The included articles were read in detail to establish the full concept and map the research on the topic of ML applications for oil and gas pipeline failure prediction. In order to facilitate the understanding of this SLR, reviewed studies were categorized into two categories including review articles and studies on the predictive models based on ML techniques which in turn classified into artificial neural network (ANN) based studies, support vector machine (SVM) based studies, Hybrid machine learning based studies and other studies that applied other ML techniques. Review articles on this topic, particularly on specific oil and gas pipeline defects, form the first part of this SLR and contributed 9.1%, which could be an adequate body of literature but because each review discussed the topic from a different and specific perspective, still there is a need for a comprehensive SLR to combine all researchers efforts in one article. Applications of different ML techniques for predicting

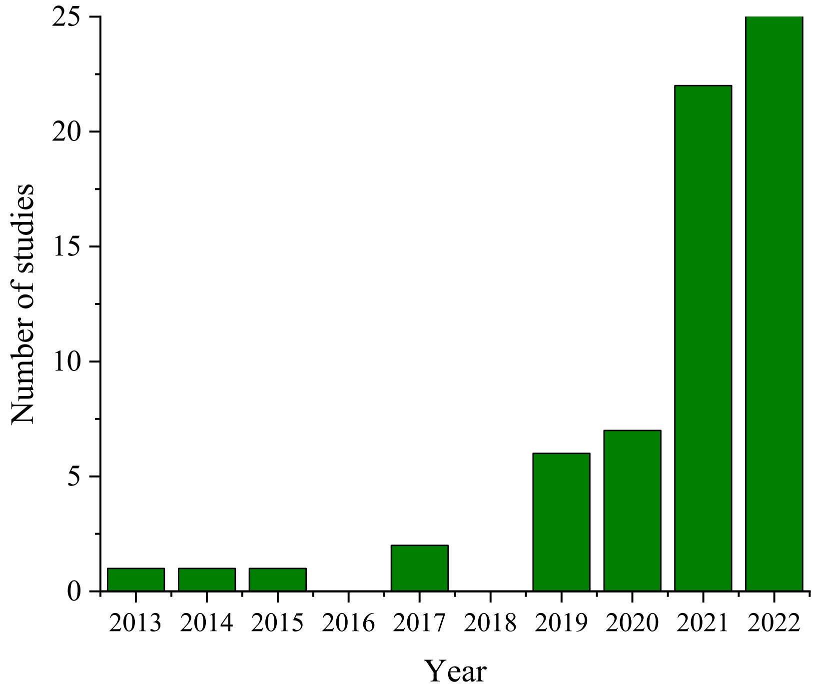

Fig. 1. Application of ML in oil and gas pipelines failures prediction trend over time.

different oil and gas pipeline failures form the second and largest portion of the included articles. The first part of this portion included 29.55% that applied artificial neural networks to develop predictive models for various pipeline defects. The second part included 27.25% that combined various machine learning techniques together such as ANN and SVM and others called hybrid machine learning (HML) techniques to compare the applicability of techniques and study the cross effects to develop predictive models for different pipeline defects. The third part included only 4.55%, which is considered the smallest body in the articles included in this SLR, that focused on using support vector machines for developing the predictive models for pipeline failures. In the last part 29.55% conducted studies on using machine learning techniques rather than the techniques aforementioned.

3. Results and analysis

The results of this SLR are presented in relation to the ML technique used for predicting the oil and gas pipeline failure in terms of the overview of oil and gas pipeline failures and the applications of artificial neural networks, support vector machines and hybrid machine learning models on the prediction of oil and gas pipelines failures. Besides the summarizing of the previous review articles that address the applications of ML in pipeline defect predictions. Overall, the trend of studies on the applications of ML techniques for oil and gas pipelines failures prediction is presented in Fig. 1. The figure clearly indicates a notable rise in the number of studies conducted in the field of ML techniques for oil and gas pipeline prediction. The highest number of studies was observed in 2022, and this upward trend began in 2019 and continued until 2022. The increase in studies during this period can be attributed to advancements in ML algorithms, a growing recognition of the importance of pipeline integrity, and the availability of comprehensive pipeline data. These factors have enabled researchers to develop accurate prediction models for preventing failures. This trend reflects the recognition of the benefits of using advanced data analysis methods in pipeline integrity management, emphasizing the need for ongoing research to enhance prediction models. Besides, Fig. 2 exhibits that several countries worldwide are interested in the applications of ML techniques for oil and gas pipeline prediction technology. China, Canada and the USA are the three major countries interested in this research field. This interest might be due to the huge oil and gas pipeline networks in these countries that need advanced technology to mitigate the failures and avoid the effects on the environment and economy. In addition, other countries such as India, the UK, Malaysia, etc. started applying this ML technology and this research pathway is still a hot topic and needs to be further investigated to come up with sustainable solutions that can mitigate pipeline defects by early prediction and reserve the environment from the oil and gas pipeline leakage pollution. The global interest in ML techniques for oil and gas pipeline prediction highlights the importance of this research in ensuring pipeline integrity and safety. Ongoing research and development are crucial to enhance the accuracy and reliability of prediction models and establish sustainable practices for long-term pipeline integrity worldwide.

3.1. Overview on oil and gas pipelines failures

Pipelines leakage is a result of many reasons such as product errors, corrosion, fluctuation in pressure, external factors and etc. Based on the literature, leakage can be diagnosed and detected based on various parameters including the pressure, mass flow rate, the size and location of the leakage and the time needed to identify the leakage to avoid more loss. Thus, such issues need to be detected early to avoid the loss of products and harm to the environment (Dai et al., 2017; Vandrangi et al., 2022). Oil & Gas (OLGA) is one of the commonly used software to monitor leak localization and sizes. It gives quick results by handling a good amount of data. Using OLGA software can easily detect large, moderate and small leaks, however, one of its disadvantages is that cannot be used to conduct a comprehensive investigation of precise points on the pipelines (Vandrangi et al., 2022).

It can be said that; the inspection and monitoring of oil and gas pipelines is facing several challenges which have to take into consideration. During the pipeline design stage and the issues and limits faced by the designers can be the first challenge. For

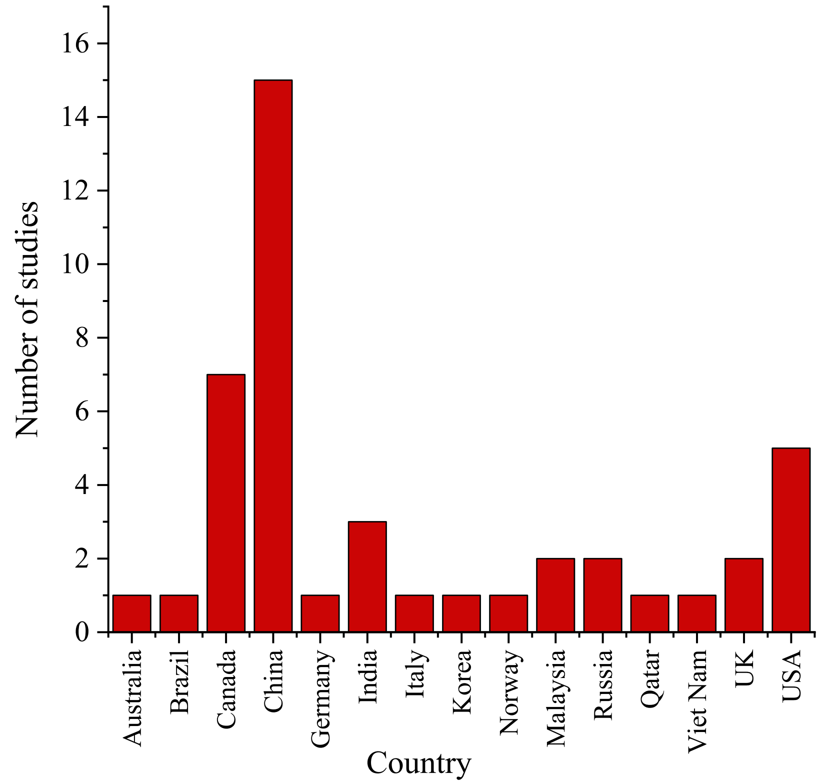

Fig. 2. Application of ML in oil and gas pipelines failures prediction versus countries.

example, the available budgets, the nature of the lands where the pipeline should be located to provide adequate protection, the raw and selected materials, the thickness of the pipeline wall and the environmental conditions around the pipeline. These all factors have direct effects on pipeline inspection and monitoring (Ho et al., 2020; Vandrangi et al., 2022). Besides, the raw materials and welding defects are considered another challenge. Even the issues in the process of pipeline laying can introduce excessive bending stress which results in the bucking and fractures. Furthermore, the offshore pipeline can expose to vortex-induced vibrations which can vibrate the length free spanning of the pipeline (Liu et al., 2014). It can be said that in order to predict and monitor pipeline failures onshore and offshore, there are several expected and unexpected challenges besides the aforementioned to be taken into consideration by researchers, industries and governments to ensure the safe and adequate oil and gas flow through the pipeline network.

To reduce the risk levels of oil and gas pipeline failures and maintain them within an acceptable range, different technologies have been employed. During the last two decades, inline inspection (ILI) has been known as the best available technique in pipeline inspection (Khan et al., 2021). This technique improved the deeper understanding of failure mechanisms, cost-effectiveness and minimize the uncertainties due to missing information. Besides, the application of the ILI method in terms of quantitative risk assessment comes up with valuable data that can be successfully used for developing accurate prediction models on the failure progress (Khan et al., 2021; Xie and Tian, 2018). However, ILI tools that are used for assessing the failure of pipelines are like other instruments affected by systematic and random errors. The systematic error type is known due to the accuracy of the instrument, while the random error is associated with the environmental conditions and parameters (Al-Amin et al., 2012; Khan et al., 2021; Wang et al., 2015). Therefore, many studies have been conducted during the last few years to address and minimize such errors.

In general, oil and gas pipeline failures can be categorized into five categories as shown in Fig. 3. It can be observed that the third-party defect contributes to 33% of total failures which is considered the highest among the other failure reasons which is the failure or loss that is due to the third party's activity. Corrosion comes after the third party which contributes to 30% including internal and external corrosion. The failures due to the design or materials properties (mechanical failures) contribute to 25% which consider a high percentage that should be taken into consideration for further research to mitigate such defects in the future. Operational, natural and other failures contribute to around 7.5%, 4.5% and 1%, respectively, which are considered lower common defects or have minor effects if they are compared with the first three failure types aforementioned. Understanding the various types of pipeline failures is crucial for keeping pipelines safe and secure. By focusing research on the main causes of failures like third-party defects, corrosion, and mechanical failures, we can create effective strategies to prevent and minimize these problems. It is also important to study and address the less common failure types to prevent them from causing major issues. Taking this comprehensive approach will ensure that oil and gas pipelines remain safe and reliable in the long run.

Corrosion defect is one of the oils and gas pipeline defects that received an intensive evaluation and investigation in the literature. Therefore, some relevant studies on the corrosion defect are summarized in this review.

The main reason for corrosion is the interaction of pipelines with different chemical components of oil and gas materials (Wasim and Djukic, 2022). Corrosion is one of the most well-known hazardous damage in oil and gas pipelines. According to the World Corrosion Organization (Koch et al., 2016), around 2.5 trillion USD is the cost of the damage caused by corrosion worldwide. That is a result of the consequences of corrosion such as environmental pollution, financial loss and heavy casualties. It

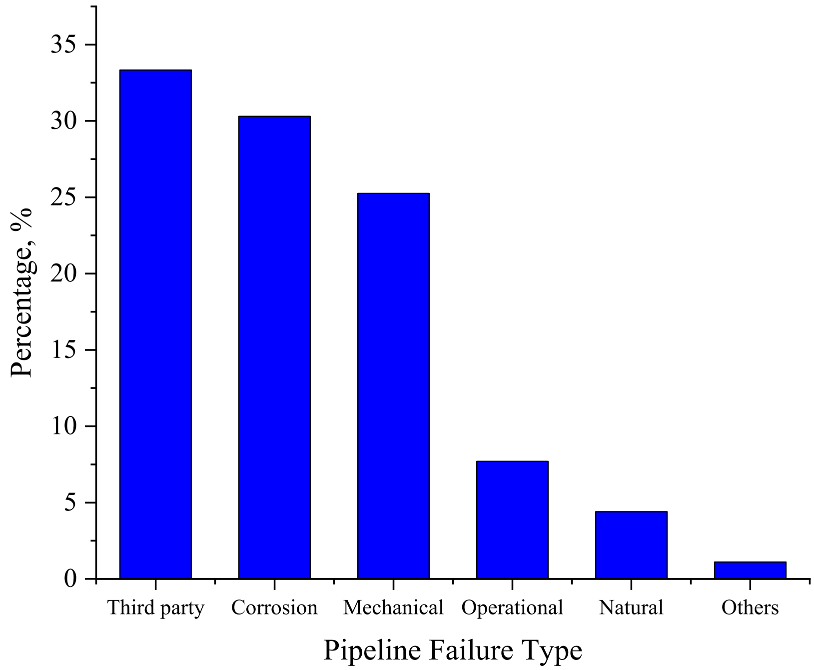

Fig. 3. Types of oil and gas pipeline failures (Soomro et al., 2022a; Zakikhani et al., 2020b).

was reported that corrosion is one of the main pipeline failures which is accounting for about 30% of the overall pipeline failures (Liu and Bao, 2022). That is the main reason for the extensive corrosion studies that have been conducted. In contrast to other failure types which still need more evaluations (Zakikhani et al., 2020a).

Different corrosion types of oil and gas pipelines were reported in the literature as shown in Fig. 4. In order to understand the mechanism of corrosion in oil and gas pipelines, various types can be illustrated as follows:

The most common corrosion that could occur in oil and gas pipeline is when the gas flow from the reservoir at high pressure and temperature which could lead to failures. Such failure is also depending on the presence and the reaction of CO_2 with the metal surfaces of the pipelines (Soomro et al., 2022a; Wei et al., 2022). Another corrosion type is pitting corrosion which is occurring due to the CO_2 presence in the pipe as a result of low velocities of oil and gas. It was also indicated that the presence of chloride ions is another main reason for pitting corrosion. It can be also stated that due to the stress generated during the oil and gas flow and the effect of flow disturbances, the inside film of the pipe is deteriorated and leads to high-intensity corrosion. It was also stated that the sources of the oil and gas can play a role in the corrosion of pipelines due to the different chemical components and interactions among the oil and gas from different resources and the film of pipelines from inside.

Stress corrosion cracking (SCC) is also another common corrosion type in oil and gas pipelines. It mainly depends on the pipe materials' sensitivity to stress, the organic solvent used and the applied tensile stress. This corrosion type occasionally leads to the failures of pipelines due to the spreading of cracks (Adegboye et al., 2019; Soomro et al., 2022a). It is crucial to address and mitigate the various types of corrosion in order to ensure the integrity and reliability of oil and gas pipelines. By understanding and taking action against corrosion caused by factors such as CO_2 presence, pitting corrosion, and stress corrosion cracking, we can prevent failures and maintain the overall safety of the pipelines. Implementing effective corrosion control measures is essential for the long-term operation and performance of oil and gas pipelines. Several factors control the appearance of such failure which were summarized in Fig. 5.

Erosion is also one of the most common issues commonly found in pipelines due to the chemical interactions among the fluid product, solid particles in the pipeline and the surrounding materials. Erosion may cause a deterioration in the pipeline materials that leads to failure. This failure may lead to serious issues for the surrounding environment, health and property. Therefore, it was recommended that the early detection of erosion rate is essential to maintain safe, cost-effective and sustainable operation conditions (Liu et al., 2021). Erosion often occurs due to the colliding of solid particles with the pipe wall during the transport of oil and gas products. That leads to gradually removing metal at the inner surface of the pipe leading to erosion degradation (Shaik et al., 2022). Erosion mainly causes the essential pipeline wall thickness reduction and endangers the efficiency of the pipeline to resist the applied pressures that leads to failures (Hu et al., 2011; Liu et al., 2021). To maintain the integrity of pipelines, it is crucial to address erosion. Regular inspections, monitoring erosion rates, and using erosion-resistant materials or coatings can help detect and mitigate erosion. Taking proactive measures to manage erosion helps prevent failures, ensure efficient operations, and minimize risks to the environment and surrounding communities.

In order to contribute to solving the erosion issue in the oil and gas pipelines, many researchers and pipeline investigators have collected data and developed models and most of the models reported in the literature combined mechanistic, empirical and computation-based methods. That combination was due to the complication of the erosion mechanisms as a result of its involves interaction among oil and gas products properties, pipelines materials, particles characteristics, pipelines geometry, operation conditions and flow regime (Karimi et al., 2017).

Several traditional techniques are used for integrity assessments of corrosion failures of oil and gas pipelines such as non-destructive evaluation methods, fault detection techniques, pipeline failure pressure methods, inline inspection methods, failure prediction techniques and burst pressure and structural integrity assessment. However, the applications of machine learning models are considered a new approach that has been used for

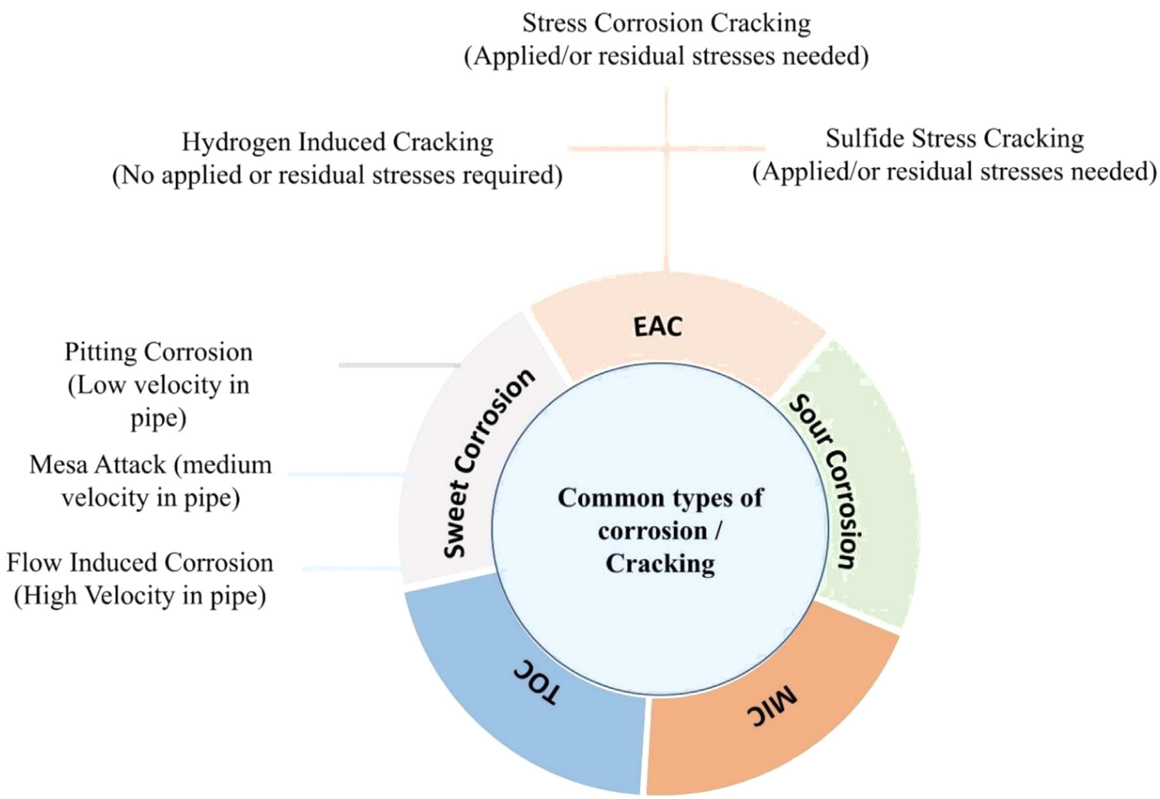

Fig. 4. Corrosion failure types in oil and gas pipeline (Soomro et al., 2022a).

corrosion failure and other failure types of oil and gas pipeline assessments and prediction which results in higher accuracy compared to the traditional techniques aforementioned. Therefore, in this review, a comprehensive systematic review is conducted to highlight the efforts of researchers in this research area.

3.2. Diagnosis and detection of oil and gas pipelines failures

The diagnosis and identify the real errors or defects along the pipeline are one of the major challenges. Besides, establishing the correct and specific causes behind the diagnosed problem or fault. Then the details and magnitude of the fault and compare it to the fault-tolerant standard is established. The correct diagnosis and collect data are the main and most important steps toward developing efficient detection and prediction systems, where detection is only the process through which the presence of any fault in a certain system is determined based on the diagnosis outputs (Vandrangi et al., 2022; Willersrud et al., 2015).

Most of the time the collected data from the industries need to be cleaned and filtered before it can be used for running and developing models to represent the actual behavior of data. Fig. 6 shows the flowchart that has been used by Khan et al. (2021) to clean the datasets obtained from the consecutive in-line inspection (ILI) runs reporting for more than seven years of corrosion failure of more than 200 km across the Canadian pipeline. The mentioned flowchart can be useful to be applied for preprocessing of data for conducting analysis and developing models to reflect the real situation of various pipeline failures. In their study, the stochastic models were presented and the variables required to model time-dependent structural integrity such as corrosion, burst pressure and containment failures were defined. Besides that, the large filtered and clean data provided by this study could be an essential source for pipeline failure assessments. Using the flowchart in Fig. 6, industry professionals can clean and filter collected data effectively. This step ensures that the data is suitable for running and developing models that accurately represent real-world behavior. Preprocessing the data in this way is crucial for conducting thorough analyses and developing reliable models to assess and address pipeline failures.

Different techniques and methods have been applied to predict the sizes and locations of leaks in pipeline networks. These techniques can be categorized into hardware-based techniques and software-based techniques. In hardware-based techniques which are rarely used these days, precise instruments are used to detect the leak from outside of the pipes which is considered an expensive technique. On the other hand, In the software-based techniques which are the most commonly utilized, continuous software analysis-based programs are used to monitor and check the flow rates, pressures, temperatures and/or other pipelines parameters based on the network of sensors installed along the pipeline (Vandrangi et al., 2022). It can be also said that several techniques of leakage detection systems have been used, whether interior or exterior. All used techniques show their strengths and weaknesses. The leakage detection systems were categorized in the literature into data-based, experience-based and model-based methods. The methods that can be selected is depended on the required accuracy, the complexity of the method, the amount of training data that is required and the cost of installations. Fig. 7 exhibits the rank of various methods used based on the complexity, amount of data required and accuracy. From Fig. 7, it can be noted that machine learning approaches such as ANN, CNN, KNN, SVM and Fuzzy showed a low complexity and high to very high accuracy to be used for pipeline defect detection, however, a high amount of data is required. That makes it one of the stronger techniques that can be applied for the prediction of oil and gas pipelines failures and defects. It is important to emphasize that the quality and quantity of available data play a crucial role in developing accurate models for pipeline failure detection using the three aforementioned methods. Adequate and reliable data is essential for training and validating machine

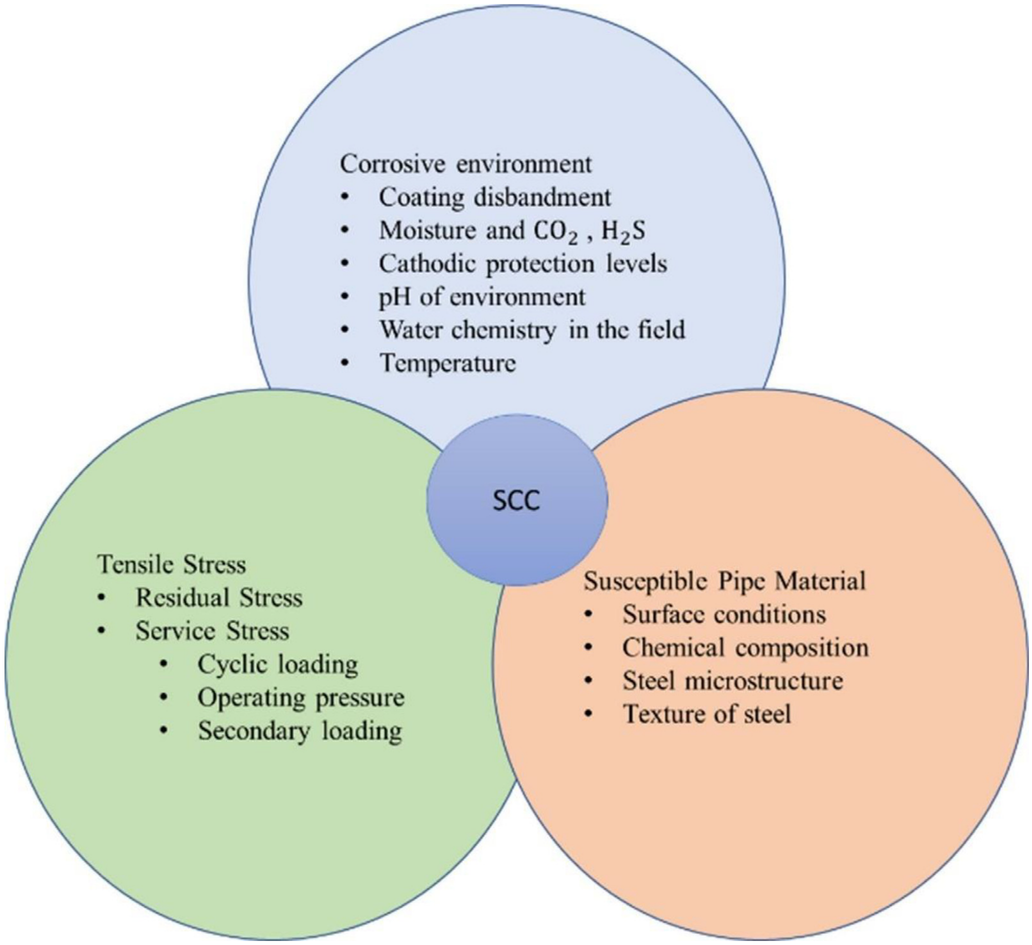

Fig. 5. The most common factors leading to initiate the SCC in oil and gas pipeline (Mohtadi-Bonab, 2019; Soomro et al., 2022a).

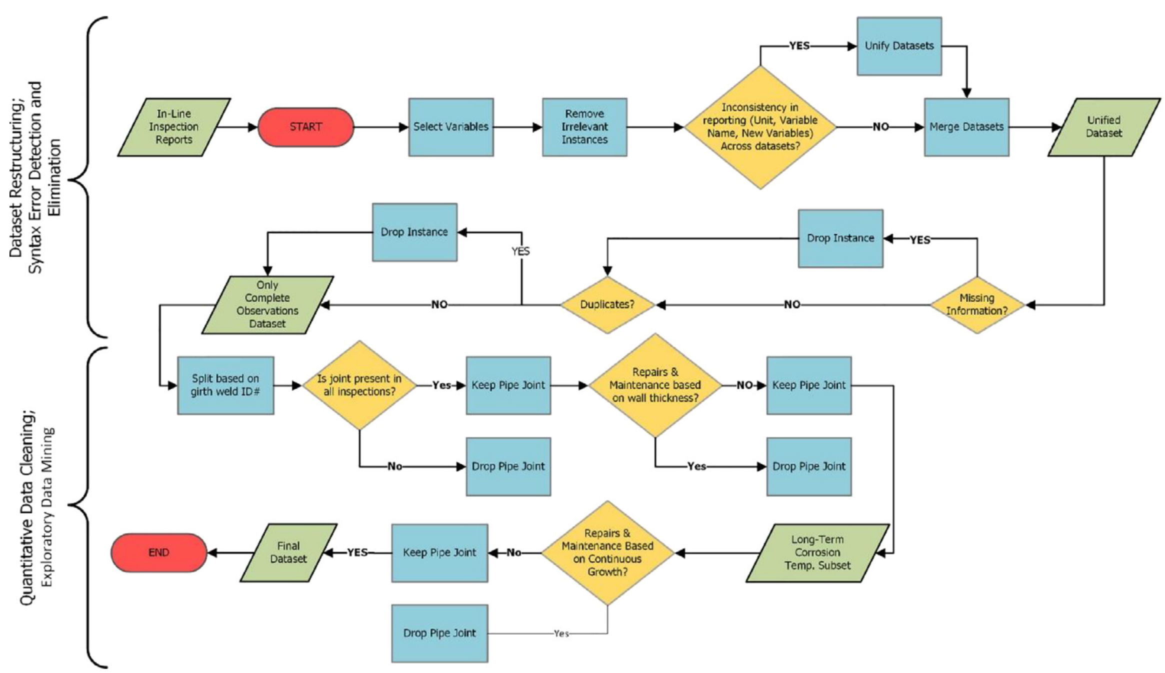

Fig. 6. Flowchart of cleaning datasets prior to being used for modeling (Khan et al., 2021).

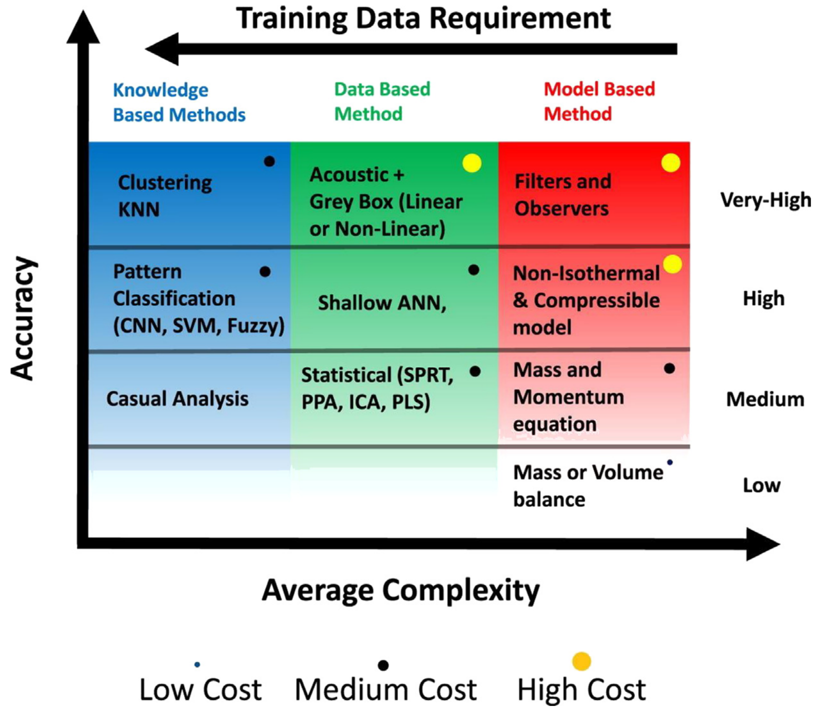

Fig. 7. Accuracy and complexity of oil and gas pipelines leak detection techniques (Vandrangi et al., 2022).

learning algorithms, ensuring their effectiveness in detecting and predicting pipeline failures. Overall, Fig. 7 highlights the importance of selecting appropriate methods based on factors such as accuracy, complexity, data requirements, and cost. Machine learning techniques demonstrate promising potential, but they rely heavily on the availability of sufficient data. By considering these factors and leveraging high-quality data, industry professionals can develop robust models for accurate pipeline failure detection and prevention.

According to the recent literature, three types of fault diagnosis techniques were reported which are data-based, model-based and experience-based or knowledge-based. Artificial neural networks (ANNs) are one of the data-driven based models that have been recently used for developing predictive models that can predict the defects and failures of oil and gas pipeline networks. It can be stated that ANNs are also one of the well-known and well-framed ML techniques that simulate the functioning of the human brain to capture the dynamics of complex behaviors and functions. By inputting some sets of data into ANN, it can identify the data patterns (Sukarno et al., 2007). Santoso et al. (2014) assessed the effect of the pressure difference and flow rate on leak detection in the pipeline based on the ANN technique. It was found that the composite effect of pressure and flow rate showed a significant effect to identify the presence of leaks in pipelines, in contrast to the separate effect of both factors that cannot show the leak.

Support vector machines (SVM) technique is another well-known ML technique that is used to distinguish the data into categories based on their features. It is also one of the recent technologies used for detecting the failure of oil and gas pipelines. To develop a predictive model for pipeline failures, the SVM is trained by using the data collected from the experimental or simulation studies such as data for leaks at different positions and various sizes along the pipeline. For non-linear analysis, there are several available kernel functions that can be used to classify the input data. Some of that functions are polynomial, sigmoid, gaussian and hyperbolic (Vandrangi et al., 2022).

To develop a predictive model, the first and most important step is to decide on the sources and the quality of the data that will be used for such modeling. The quality of data is mainly depending on the accuracy of the instrument, tools and other techniques that have been used for data collection. However, data cleaning is one of the most important factors that contribute to the highly accurate model. For example, noise data and irrelevant data are deleted from the datasets that will be used for developing the model. There are two main deletion approaches have been used in literature for the same purpose list wise and pairwise (Little and Rubin, 2019). In list wise method, samples with missing information on any variable are excluded. In pairwise technique, the data is removed if the missing variable is used in the analysis (Roth, 1994). After the data cleaning, categorizing the data based on their features is performed using the summary statistics and shape of the distribution to ensure the quality of the data (Witte and Witte, 2017).

Applications of ML techniques in oil and gas pipelines have been attracting research during the last few years. That could be due to the ability of ML to provide higher accuracy, lower cost and time compared to conventional methods. Artificial neural networks, support vector machines and decision trees are the common machine leering approaches that have been studied to develop the predictive models of oil and gas pipeline failures as shown in Fig. 8. According to VOSviewer mapping and density visualization, the deep neural networks technique has not received enough attention to be applied to oil and gas pipeline defects. It can be also noted that corrosion failure is the most common failure that has been investigated. In contrast to other pipeline failures which received minor attention and further studies are required to apply different ML techniques to develop predictive models for different pipeline failures such as weldment defects, materials defects, natural defects and so on. It can be also noticed that nondestructive examination and magnetic flux leakage tests are the most common tests used for generating the laboratory data, which indicates further up-to-date and more reliable tests and advanced simulations such as finite elements methods could be further involved to generate data required for more accurate ML models. From Fig. 8, it can be summarized that, while ML techniques have shown promise

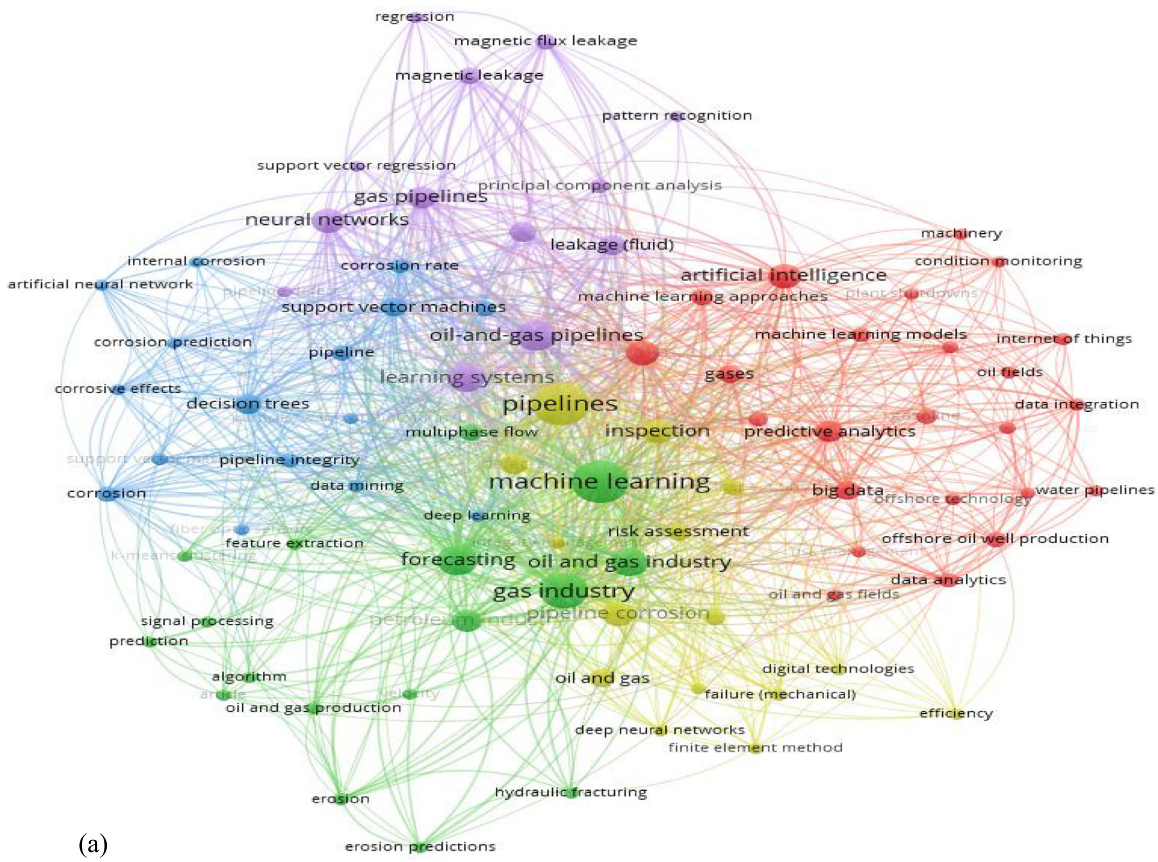

Fig. 8. VOSviewer of authors' keyword analysis sources with a minimum of four occurrences in an article (a) Mapping (b) Density visualization.

in the oil and gas pipeline industry, there is a need for further exploration of deep neural networks and the application of ML models to different types of pipeline failures. Additionally, incorporating more advanced testing methods and simulations can enhance the accuracy and reliability of ML models in predicting pipeline failures. Fig. 9 highlights the global adoption of ML techniques for predicting oil and gas pipeline defects. China, Canada, and the United States lead this research due to their extensive pipeline networks, requiring advanced technology to prevent failures and mitigate environmental and economic consequences. However, further exploration is needed in other countries to develop sustainable solutions. Expanding research efforts worldwide is crucial for comprehensive approaches that address region-specific challenges.

3.3. Studies on ML applications for oil and gas pipelines failure assessment

Fig. 10 shows the distribution of the studies included in this SLR for the four main categories of reviews, studies applied ANN, studies applied SVM, studies applied HML and studies applied other ML techniques. It can be seen that reviews account for 9.09% of the studies, indicating a smaller focus on literature synthesis. Studies using ANN contribute the highest percentage at 29.55%, indicating its widespread use for pipeline defect forecasting due to its ability to handle complex relationships. SVM studies contribute 4.55%, indicating its effectiveness for high-dimensional data and limited datasets. HML studies contribute 27.27%, showing the exploration of combining ML techniques for improved predictions. Additionally, studies using other ML techniques contribute 29.55%, revealing a variety of ML approaches being experimented with. Overall, Fig. 10 demonstrates the diverse range of ML techniques employed in oil and gas pipeline defects prediction.

In the following sections, a comprehensive review of ML applications for oil and gas pipeline failure assessment is provided. Existing literature reviews on the use of ML techniques and their key contributions are discussed. Specific ML approaches, including artificial neural networks, support vector machines, and hybrid models, are explored, and an overview of relevant studies and their applications in predicting pipeline failures is presented. Our goal is to offer a comprehensive understanding of the current state of ML in this important area of pipeline defect detection.

3.3.1. Reviews on ML applications for oil and gas pipeline failure assessment

Surprisingly, no more review articles have been found in the literature on the applications of ML for oil and gas pipeline failure prediction. However, such new technologies should be introduced to spread the knowledge to other interested researchers, industries and communities. Furthermore, ML approaches from different perspectives and descriptive statistics should be established to explain the different ML techniques and models for improving the detection of various oil and gas pipeline defects and failures. Rachman et al. (2021) have reviewed the applications of ML in oil and gas pipelines from the pipelines integrity management (PIM) perspective. They mainly focused on the application of ML in processing and managing data generated from PIM activities. PIM activates or processes are included inspections, monitoring and maintenance. Types of inputs, outputs, pre-processing and learning algorithms of ML that have been applied in different PIM stages were discussed. Besides, the research challenges, future directions on the applications of ML

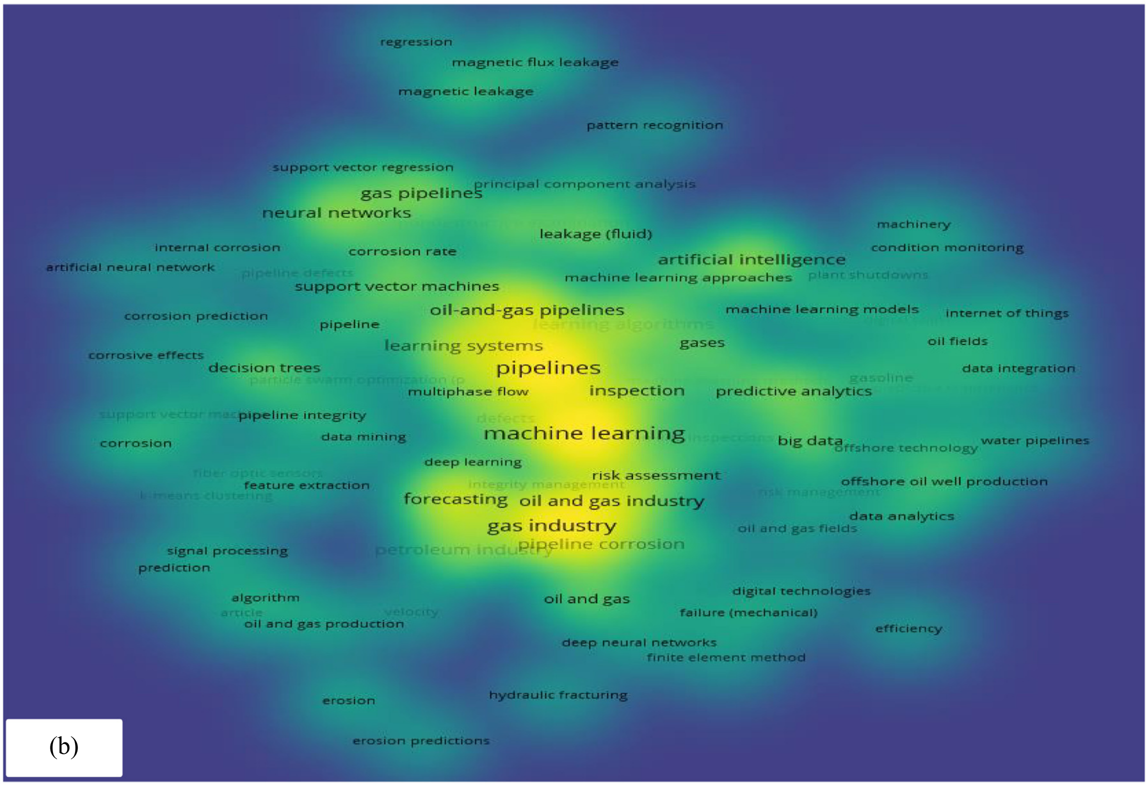

Fig. 8. (*continued*).

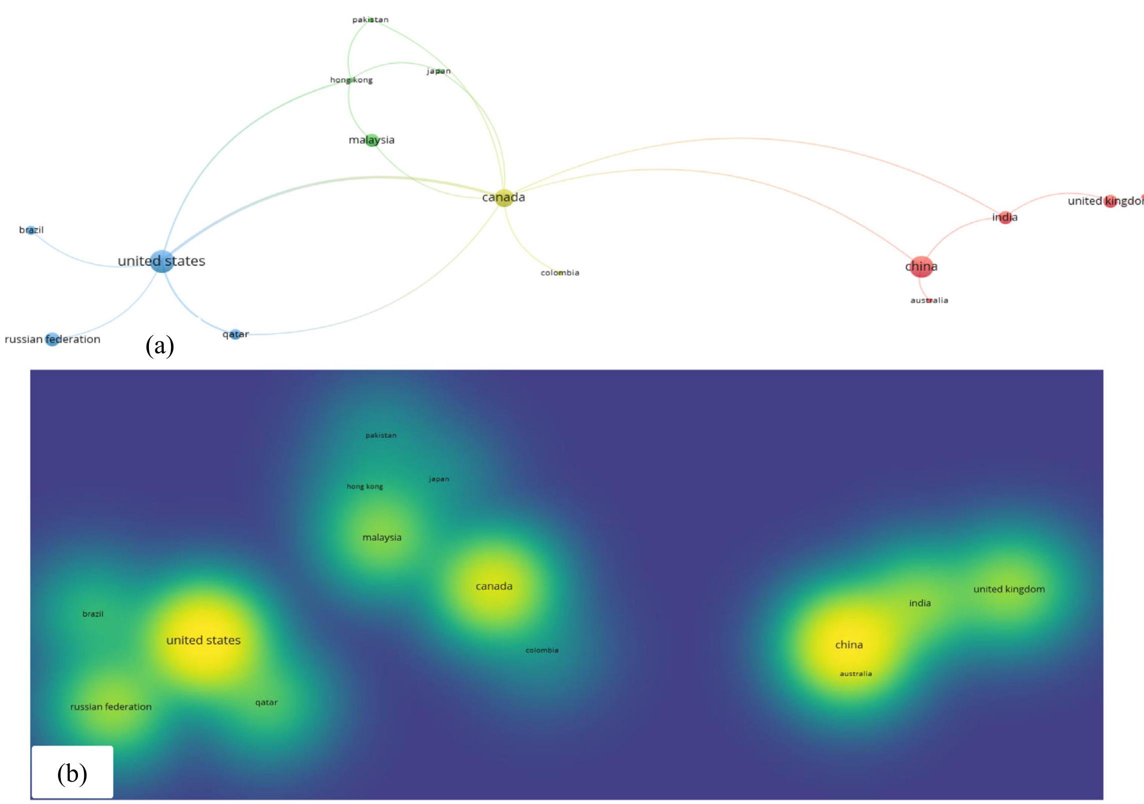

Fig. 9. VOSviewer mapping and density visualization of countries with at least one article published (a) Mapping (b) Density visualization.

for PIM of oil and gas have also been introduced. Another article introduced leakage detection techniques for pressure failure of pipelines (Zaman et al., 2020). However, this article mainly focused on the mathematical approaches and hydrological tools for pressure defect prediction. Therefore, it is quite far from our concern for this SLR. Soomro et al. (2022a) systematically

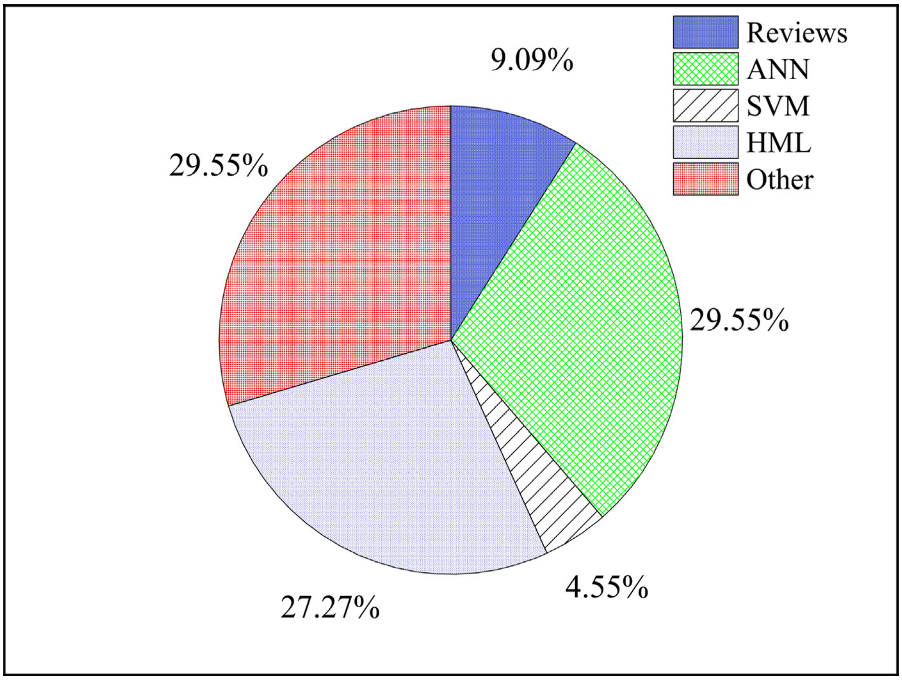

Fig. 10. Distributions of the included studies on five main research categories.

reviewed the applications of ML in corrosion failures of oil and gas pipeline predictions. The different ML techniques including the ANN, SVM and HML on the prediction of corrosion failure of pipelines have been summarized and discussed. The variables have been identified by literature and dataset sources used to generate the data for developing ML models for corrosion defects have been also comprehensively addressed. It was reported that HML techniques showed to be a significant ML technique compared to other standalone ML models for higher accuracy of prediction. It was also stated that out of the experimental, simulation and the field data sources used in the literature, field data source is most commonly used. Some future recommendations and suggestions were also proposed for further exploration in the ML applications for corroded oil and gas pipeline. In their recent review, Soomro et al. (2022b) examined the existing applications of the Bayesian network approach in the detection of corroded oil and gas pipelines. The findings indicated that a majority of the research studies utilized Bayesian models with insufficient data, resulting in unforeseen outcomes.

Therefore, it can be said that few review articles, especially systematic reviews, have been reported that introduce the application of ML in oil and gas pipeline failure predictions. That results in a wide knowledge gap in these technologies. Fig. 11 shows the summary of the number of reviews has already been published, besides the gaps that need additional reviews on ML applications for oil and gas pipeline failure prediction technology. The highest number of reviews has been reported for the applications of ML in corroded pipeline detection, followed by ML for integrity management and HML for pressurized pipelines. However, researchers should publish SLR based on different perspectives such as on ML for pipeline failures (including a comprehensive and connected mapping for the application of ML for different pipeline failures), ANN for pipeline failures, SVM for pipeline failures, HML for pipeline failures, ML for mechanical failures of the pipeline, ML for third party failures of the pipeline, ML for operational failures of pipeline and ML for natural failures of the pipeline. Furthermore, systematic literature reviews from the civil engineering researcher's perspective, from petroleum engineering researcher's perspective, from mechanical engineering researcher's perspective and from data and information technology scholar's perspective are recommended which could be very useful for a comprehensive overview of the applications of ML in oil and gas pipelines failure prediction. In summary, the lack of systematic reviews on ML applications in oil and gas pipeline failure predictions has created a knowledge gap. Additional reviews are needed to cover various areas, including different types of pipeline failures, ML techniques, and perspectives from different engineering disciplines and data and information technology scholars, to provide a comprehensive overview.

3.3.2. ML approaches for oil and gas pipelines failure assessment

In this section, the studies that have been conducted in the literature are classified based on the most common ML models (artificial neural networks, support vector machines and hybrid machine learning models).

3.3.2.1 Studies on artificial neural networks applications for oil and gas pipeline failure prediction The artificial neural network is one of the wide ML techniques that have been employed during the last few years to develop predictive models that can accurately predict the various defects of oil and gas pipelines. In this regard some of the studies conducted were summarized in this subsection considering the type of ANN used, parameters and responses considered, the source of the data used for training the models and the main findings.

Mohamed et al. (2015) studied the applications of ANNs at various architectures to predict the pipeline defect depth. Data obtained from the pipeline operators using magnetic flux leakage (MFL) sensors have been used in this research. Prior to using the Levenberg–Marquardt backpropagation learning algorithm to train the model, different defect depths and features have been extracted. The magnitude of MFL signals used as input in this study to establish the relationship with defect depths. The three architectures that have been investigated are static feed-forward neural network (FFNN), Cascaded FFNN and dynamic FFNN. It was reported that LM back-propagation showed to be the best learning algorithm compared to other learning functions for defect depth prediction. It was also found that dynamic NN yields the

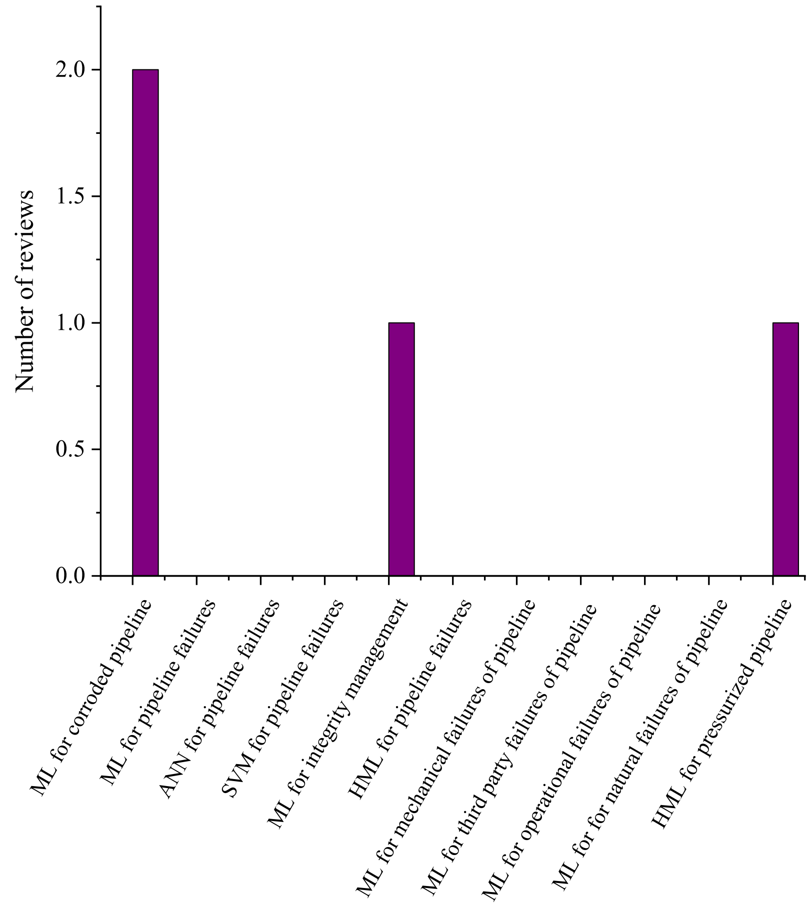

Fig. 11. Review articles on ML applications for oil and gas pipelines failures prediction.

best performance of 89% accuracy and cascaded NN yields the worst performance among the three NNs used. A novel real-time action recognition technique has been developed using CNN to enhance the accuracy of automated pipeline safety early warning (PSEW) systems that are commonly used for identifying and locating third-party defects on oil and gas pipelines (Yang et al., 2021c). The data was collected from the optical fiber sensors along the pipelines at the site at China National Petroleum Corporation Pipeline in 2016 along the 48 km pipeline. Manual excavation, vehicle driving over the pipeline, and mechanical excavation parameters that cause the pipeline defect have been identified and considered. The developed model showed to fulfill the industrial requirements to be applied with an adequate degree of accuracy.

Ossai (2020) carried out research to develop machine learning models to predict the corrosion depth defect and burst pressure of oil and gas-aged pipelines. Subspace clustered neural networks (SCNNs) have been used to optimize the weights at the clusters of feed-forward neural networks to improve the accuracy of defect depth prediction. The mechanical parameter including (diameter, thickness, length, yield strength and ultimate strength of tested pipelines) and operating parameters including (pressure, temperature, gas production rates, oil production rates, water production rates, specific gravity and CO_2 partial pressure) were considered for developing the predictive model. The model was trained with onshore pipeline data. It was stated that the developed model could provide a guide for experts on the defect depth prediction over time, and provide baseline information for effective management of the operating pipelines toward lower cost and safe production environment. In order to mitigate the aggregation of hydrate in the gas pipelines that cause a serious problem for the flow and stability of production, Seo et al. (2021) established a system based on the feed-forward ANN to predict the hydrate in the gas pipelines from flow assurance perspective. The model was trained with data obtained from the OLGA simulator and parameters of pressures, temperatures, and hydrate volumes at each time step. It was reported that ML can be a useful technique to predict the hydrate information in real-time with adequate accuracy.

The deep neural network technique was used to develop fast and accurate predictive models for the failure pressure of oil and gas pipelines (Su et al., 2021). 142 groups of data obtained from the burst pressure test, besides 150 groups of data generated from FEM simulation have been used for DNN model training and validation. The effects of different defect size on pressure failure was investigated extensively. It was found that the applications of multilayer DNN on the modeling of failure pressure can provide a high efficiency and satisfactory accuracy which can be generalized and applied for the assessment of oil and gas pipeline integrity. Zakikhani et al. (2020b) developed predictive models based on ANN to predict the failure sources of oil pipelines in order to overcome the disadvantages and inefficiencies of the several inspection methods that have been used previously to predict the failure of oil pipelines which results in economic losses and environmental impacts. Data used for developing a model was obtained from a field record for 35 000 km of pipelines over 37 years across Europe. For predicting the mechanical, corrosion and third-party failures as outcomes, pipelines age, service types, diameter, land use and facility types were used as input parameters. It was found that the developed models could be useful for predicting the mechanical, corrosion and third-party failure of oil and gas pipelines with an accuracy of around 72%.

That will help the decision-makers and pipeline operators to prioritize inspections and provide an adequate view of the failure sources allowing decision-makers to take a decision to mitigate the risks. It was also stated that pipeline age and product type showed the least and most impact on the failure source.

Ferreira et al. (2021) applied the deep neural networks technique to develop a predictive model that can be used to predict burst pressure failure in pipelines. Experimental data obtained from Ultrasonic inspection for the corrosion dimensions in terms of length and thickness of the defect were used as input for finite element analysis. Then DNN was trained with other data obtained from finite element analysis. Developed DNN models showed a high degree of accuracy to predict the burst pressure failures of oil and gas pipelines with an error of less than 2% and a coefficient of correlations of 99%. To develop a predictive model for erosion severity of oil and gas pipelines under high internal pressure, Yang et al. (2021a) employed ANN that was trained with experimental data obtained from the erosion tests involving tensile stress. It was stated that utilizing ANN provided a potential model that can be used for predicting the erosion severity of pipelines under high internal pressures with an adequate degree of accuracy. Yang et al. (2021d) conducted a study to develop a model that can be used for accurately identifying and locating the third-party failure of oil and gas pipelines using multifeatured-fusion convolutional neural networks. The developed model was trained with a large amount of data collected by the China National Petroleum Corporation from the real oil and gas pipeline networks in the country. It was found that the developed model could be used with a high degree of accuracy of more than 95% to identify and locate the third-party defect under various conditions in real-time. It was also stated that the developed model could be generalized to other fields including measurement, monitoring and industry inspection.

In order to develop models that can be utilized to predict the external corrosions in oil and gas pipelines, Bastian et al. (2019) trained a deep neural network in terms of CNN with a dataset of 140 000 optical images obtained manually from inspection videos of the corroded pipelines. It was claimed that the developed model can classify and discriminate between the images of corroded pipelines and images that have similar patterns to corroded pipelines but without real corrosion. The degree of accuracy of the developed model was found to be 98.8%. In addition, a localization algorithm was proposed to identify the corrosion area in the given images with higher precision. It was stated that the developed DNN model will overcome the disadvantages such as the higher cost and interrupting the functioning of the pipelines for manual inspection and other non-vision techniques based on nondestructive evaluation methods.

Table 2 summarizes some of the studies reported in the literature on the applications of ANN and SVM individually for developing prediction models for various defect types of oil and gas pipelines. The inputs and outputs, data sources used for training the models and main findings were highlighted and briefed. It can be seen that most of the studies used ANN which could indicate the appropriateness of this ML technique for predicting the oil and gas pipeline defects with adequate accuracy. It can be also noticed that most of the existing studies investigate the use of ANN and VSM for predicting the corrosion of pipelines. However, the research gaps in the development of predictive models to predict other oil and gas pipeline failures such as three-party, mechanical, natural and operational failures still need further investigation in the future.

3.3.2.2 Studies on support vector machines applications for oil and gas pipelines failure prediction Piao et al. (2019) carried out a study to develop a fast reconstruction of the 3D defect profile using a least-square support vector machine (LS-SVM) based on data collected from magnetic flux leakage (MFL). It was found that the proposed models' accuracy and computational speed have been enhanced significantly compared to traditional methods such as backpropagation ANN which required large data to be used for the same purpose and come up with adequate accuracy. Another Study was conducted by Luo et al. (2013) to develop a model based on grey SVM to predict the corrosion rates of gas pipelines. Known corrosion inspection data of oil pipelines have been used to train the model considering pressure, deposition rate, angle, the density of the gas, density of the liquid, liquid hold-up, liquid velocity, surface tension, pH value, fluid temperatures, inner wall surface temperature, flow regime, superficial velocity of gas, thermal conductivity of gas and maximum wall shear stress as inputs parameters. It was claimed that developed models provided a new thought for risk management, risk assessment and maintenance of oil pipelines. It was also stated that developed models could be useful for integrity management and quantitative assessment for long-distance oil and gas pipelines.

3.3.2.3 Studies on hybrid ML models applications for oil and gas pipelines failure prediction Hybrid machine learning is considered a significant technique for improving the accuracy of oil and gas pipelines failures prediction due to its advantages compared to other standalone machine learning techniques. Some of these advantages can overcome the issues of overfitting the need for huge data to train the model. Besides, the optimization process can be applied to improve the prediction accuracy and by combining more than machine learning techniques the advantages of all techniques can result in a significant model to be used for the prediction of different complex pipeline defects with a high degree of accuracy.

Peng et al. (2021) carried out a study to develop a hybrid machine learning algorithm method to predict the corrosion rates of oil and gas pipelines. Principle component analysis (PCA), SVM and chaos particle swarm optimization (CPSO) techniques have been combined together to develop the proposed models. It was stated that PCA was used to screen out the main variables that influence the corrosion rates while CPSO was utilized to optimize the hyperfine parameters in SVR to enhance the prediction accuracy of the model. A total of 60 groups of data were collected from the 5.5 km submarine oil and gas pipelines in Hainan, China. Parameters used are pH, flow rate, temperature, partials CO_2 pressures, pressures, wall shaving stresses and liquid hold, while the output was the corrosion rate. It was claimed that the proposed PCA-CPSO-SVR model showed an absolute error of 0.083 which is lower than the error from SVR alone by 18.6%. It was also stated that developed models showed higher prediction accuracy compared to PCA-GA-SVR, ANN, linear regression (LR), De warred95 (OLGA) and PCA-PSO-SVR. Liu et al. (2021) conducted a study to develop machine learning models (ANN, classification tree and Bayesian network) to help in the evaluation of erosion rate and the optimal flow velocity in pipelines. Sand production pipe saver (SPPS) V5.3 and DNV GL RP-O501 models have been utilized to calculate and generate the datasets used in developing the machine learning models. Field data was also used in this study. The Monte Carlo simulation method was also used in this study. A parametric study was carried out to show the impact of each factor on the erosion rate. VBA code is also used to evaluate if erosion is within the allowable range. Hugin Expert software is also used to establish a Bayesian network to ensure uncertainty. Overall it was found that machine learning techniques can make a faster and more efficient decision and handle uncertainty very well in a probabilistic manner compared to conventional

Table 2
Summary of studies on prediction of oil and gas pipeline failures using ANN and SVM individually.

Reference	ML technique used	Input parameters	Output evaluated	Remarks
Afebu et al. (2015)	ANN	Flow rate, velocity, pressure and temperature	Detect leak locations and leak sizes	– Data obtained from OLGA simulator. – Better precision in predicting leak locations was found compared to leak size predictions.
De Masi et al. (2014)	ANN with Lavenberg–Marquadt (LM) learning algorithm and 20 hidden neurons	Elevation, slope, concavity, flow regime, hold-up, pressures, gas flow, liquid velocities, and gas velocities	Corrosion rate, metal loss and defect area	– Data obtained from the field (case study) – It was found that developed ANN models showed significant performance compared to deterministic ones that a significant drawback in accuracy.
Santoso et al. (2014)	ANN	Flow rate (time) and pressure	Detect pipeline leakage	– ANN backpropagation with three layers showed to be the optimal structure for pipeline leakage very well.
Chamkalani et al. (2013)	ANN	pH value, velocity, pressure of CO_2 and temperature	CO_2 corrosion rate	– Data collected from the experimental work – It was found that ANN can predict the CO_2 corrosion rate with very close values to the experimental findings. – The developed model was recommended to be used for corrosion of pipeline prediction considering the limitations applied.
Din et al. (2015)	ANN	Length, width, depth and orientation of the corrosion defects	Corrosion rate	– Data was obtained from In-line inspection (ILI) data – ANN was found to be an efficient technique to predict the corrosion rates based on the mentioned inputs in this study
Nayak et al. (2020)	ANN (multilayer perceptron neural network)	pH value, CO_2, temperature and velocity	CO_2 corrosion in the oil and gas pipeline	– Data obtained from experimental and computational fluid dynamic – Corrosion obtained from ANN, CFD and experiments found to be very close and similar
Bastian et al. (2019)	Deep neural network using CNN	Images dataset was collected from oil and gas pipeline	Detect the level of corrosions	– DNN showed to be very accurate to develop models that can be used for identifying the corroded regions successfully on the pipelines
Ren et al. (2012)	Back propagation ANN	– Elevation of pipeline – Pressure – Length of pipeline – Pipe slope	Internal corrosion rate	– Data obtained from an experimental program – ANN model has been developed to predict the internal corrosion rates of gas pipelines accurately.
Carvalho et al. (2006)	Nonlinear pattern classifier using ANN	MFL signals obtained from intelligent PIG	Defect in the weld zone of the pipelines (if the signals are defective or non-defective and if the defects are external or internal or non-penetrating).	– Data was obtained from the installed sensors on the 1025 points along the pipeline – ANN showed an accuracy of more than 90% for the classification of the defects – The best ANN performance found at 15 neurons in the hidden layer
Silva et al. (2007)	ANN	Pipe wall thickens, defect depth and dimensionless circumferential spicing	Gas pipeline' failure and burst pressure	– Finite Element Method (FEM) used in this study to generate the dataset – It was found that the pressure failure is associated with the length and depth of the defect
Xu et al. (2017)	ANN	Ratio of defects depth to pipe thicknesses, ratio of defects length to pipeline thicknesses, dimensional circumferential spacing and dimensional longitudinal spacing	Burst pressure failure	– Experimental data – ANN showed to be capable to predict the pressure failure from the interacting pipe defect
Chin et al. (2020)	ANN	The tensile strength of pipe, nominal thickness and nominal diameter and defect length and depth of pipe.	Normalized pipe failure pressure	– Data used from literature for full-scaled burst pressure test – It was reported that the defects depth has a proportional relationship to the pressure failure
Valizadeh et al. (2009)	Fuzzy	Flow rate, pressure and temperature of pipe	Detect pipeline leakage at different positions and sizes	– Features extractions and classifications can be applied to detect the pipeline leakage

(*continued on next page*)

Table 2 (*continued*).

Reference	ML technique used	Input parameters	Output evaluated	Remarks
Luo et al. (2013)	SVM	Pressure, deposition rate, angle, density of the gas, density of liquid, liquid hold-up, liquid velocity, surface tension, pH value, fluid temperatures, inner wall surface temperature, flow regime, superficial velocity of gas, thermal conductivity of gas and maximum wall shear stress	Corrosion rate of gas pipeline	– SVM results in adequate models that can be used for predicting corrosion rates based on the inputs in this study

methods. Developed models using machine learning provide cost-effective and user-friendly tools for erosion rate prediction to maintain safer and more sustainable pipelines. Therefore, the procedures followed in this study are strongly recommended to be used for evaluating and producing the other failure types in oil and gas pipeline networks. In order to improve the accuracy of existing engineering simulators available to predict the pressure drop and steady-state multiphase pipe flow, Kanin et al. (2019) combined several ML algorithms including SVM, ANN, Gradient Boosting algorithm and Random Forest. Data used for model training from lab data set obtained from open literature. It was found that the hybrid ML algorithms used can predict the pressure distribution along the pipes with correlation coefficients maximum of 0.99 which is considered higher accuracy and better than the mechanistic models and multiphase flow correlations used for similar prediction purposes.

To predict the erosion rate of liquid hydrocarbon pipelines toward cleaner and safer transportation, machine learning techniques have been used (Liu et al., 2021). ANN, Classification tree and Bayesian network were selected for developing the erosion rate predictive model for their simplicity and comprehensively to address the different pipeline operating sensors with adequate accuracy. Pipe diameter, the radius of curvature, steel of hardness, particle diameter, particle hardness, particle density, fluid density, fluid viscosity and fluid velocity are used as inputs for predicting the erosion rate. The dataset used was obtained from a Monte-Carlo method simulation. It was found that the classification tree technique showed to be the fastest method among the three used, ANN provides comprehensive sensitivity evaluations for each input and the Bayesian network was more useful to predict the erosion rates with uncertain data. After the comparison of three ML techniques with non-ML models, it was found that all ML techniques exhibited more efficient ways and accuracy to make a decision. Therefore, it was stated that developed models can serve as a basis for the pipeline operators to take a decision on pipeline inspections and maintenance toward minimizing the erosion rate. Yin et al. (2021) carried out a study to introduce a semi-quantitative framework that combined fuzzy logic, similarity aggregation model and ML techniques for criticality evaluations of oil and gas pipeline networks. The fuzzy logic method was used to establish the criticality index and build an easy model to facilitate the evaluation process using ML. While SAM is used for data collection strategies. It was claimed that the developed framework showed the advantages of the three techniques in terms of excellent handling of uncertainty, the reasonable aggregating of experts' opinions by SAM and the high ability of ML to fit the data. The failure scenario and affect the type of failure consequences of the oil and gas pipeline were introduced by Yin et al. (2021) and is shown in Fig. 12. Besides, the proposed framework to predict the criticality of oil and gas pipelines through the three stages of data collection, fuzzy logic inference and ML modeling that have been developed is presented in Fig. 13. The data collection strategy gathers pipeline conditions, environmental factors, and other variables influencing failures. Fuzzy logic assesses potential failure severity using collected data. Machine learning algorithms predict and analyze pipeline failure criticality. It was claimed that the framework in Fig. 13 offers a systematic approach to evaluate and predict the criticality of oil and gas pipeline failures. It uses data collection, fuzzy logic, and machine learning to assess and prioritize risks. This framework can be useful as a guideline for evaluating various failures of oil and gas pipeline at different regions around the world under different conditions.

Jiang and Dong (2020) conducted a study to provide several models based on finite element method (FEM) analysis and ML algorithms that can be used to predict the collision failure risk of pipelines. Sample space has been drawn using FEM and the Latin Hypercube Sampling technique to ensure the uncertainties and non-linear effects in the collision were considered. Four ML techniques were used in this study: Genetic programming (GP), multiple layer perceptron (MLP), SVM and radial basis function network (RBFN). Parameters considered are pipeline diameter, yield strength, wall thickness and object mass. It was reported that the GP model showed the best performance among other ML techniques used with a correlation coefficient higher than 0.999 and root relative square error (RRSE) less than 0.06. The developed model was validated by FEM showing a good correlation. Besides, the feasibility of the developed model was proved by considering nonlinear influences and pipe-soil interaction to come up with a more realistic risk evaluation. Layouni et al. (2017) conducted a current study to propose a model using machine learning to automatically analyze MFL signals to detect the size of metal-loss defects in oil and gas pipelines. They claimed that such a model will overcome the challenges of manual detection by engineers which could be cumbersome to human operators. ML techniques applied were ANN, linear and non-linear parametric regressions. ML models trained with data obtained from raw MFL signals after conducting feature extraction. Features extracted that depend on the defect's depth were maximum magnitude, mean average, peak-to-peak distance, standard deviation and integral of the normalized signal. It was claimed that developed models showed a high degree of accuracy and computational efficiency and the proposed model can detect any pattern in the MFL signal that is known as a reference pattern. As they stated the developed model can even be useful to be applied to a wide range of defects shapes that are not known because of its flexibility. Another advantage for propose model is detecting the defect sizing by the model is fully automated.

Yang et al. (2022) proposed a novel model that used a one-dimensional neural network (1DCNN) and SVM to enhance the detection accuracies of oil and gas pipeline leakages. After extraction of data features using IDCNN, an enhanced variable amplitude particle swarm optimization (VAPSO) algorithm was applied with an adjustment strategy to optimize the combination of parameters in SVM and minimize the risks of trapping. The data features extracted were then used as input into the enhanced

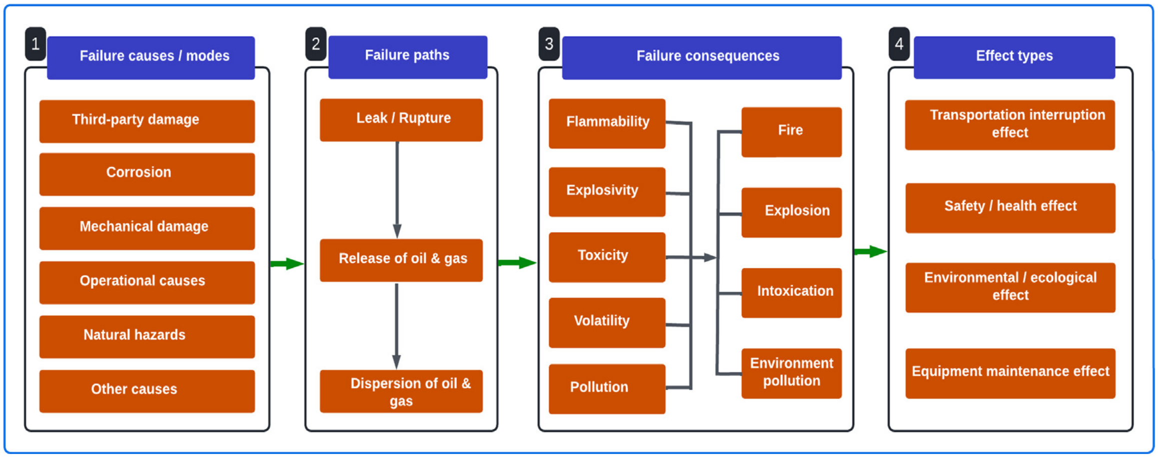

Fig. 12. Failures scenario and affect type of failures consequences of oil and gas pipeline (Yin et al., 2021).

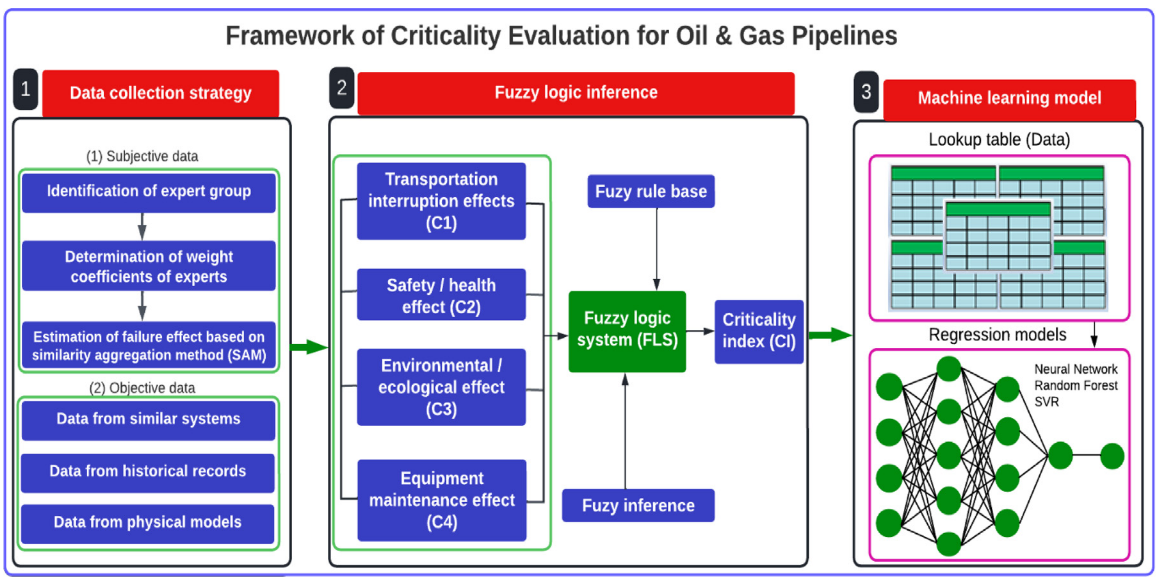

Fig. 13. Proposed framework of criticality for oil and gas pipelines (Yin et al., 2021).

VAPSO-SVM to classify. Data used in this study was obtained from the pipeline signals collected from the laboratory experiments by simulating the oil and gas pipeline with 160 m. The 8103 hydrophone sensors have been used to detect weak leakage signals in frequencies of 0.1 Hz to 180 kHz. The parameters considered are temperature, flow rate and pipeline pressure. It was claimed that the developed model showed the ability to extract the features of pipeline data more accurately and faster with effective enhancement in the classification accuracy. Overall, the developed model exhibited higher robustness in the prediction of pipelines leakages. Yang et al. (2021b) studied the effects of the tensile stresses introduced by the high internal pressures and erosion defects due to the solid particles by the fracturing fluid on oil and gas pipeline defects. Data were collected from the erosion wear experiment of fracturing pipelines in various situations. Parameters included were tensile stress, impact angle, flow velocity, target material and particle concentrations. SVM, KNN, ANN and random forest regression (RFR) models have been used to develop models that can be utilized to predict the erosion rate. RFR showed optimal agreement with the actual data obtained from the experiment with a high degree of accuracy and lower error compared to other machine learning models used.

Mazumder et al. (2021) applied several ML techniques to develop a valuable alternative model to the conventional techniques that can computationally intensive analyses and determine the failure risk of oil and gas pipelines. Data used for developing models were obtained from the experimental data that had already been published in the literature. Corrosion defect of the pipelines was detected considering the remaining strength parameters of the pipeline. Eight machine learning models were used to identify the best failure prediction model. Fig. 14 shows the flowchart of the proposed methodology adopted in the study. The XGBoost was found to be the optimal to be used for accurate failure prediction and it was recommended for future analyses. The efficiency of ML-based models was also compared to the physical-based models and it was found that ML models can perform the failures risk analysis with much better accuracy than physical-based methods. Besides, it was stated that in terms of time, ML is 12 times faster than physical-based methods. Liu et al. (2019b) conducted a study to develop an ML model that can automatically much the growth of corrosion with extracted data from In-Line inspection (ILI). After the extraction of features from ILI data, SVM, decision tree and random forest ML models have been used individually and ensembled. It was stated that accurate matching was found in the case of using the ensemble learning method. Therefore, this characterization of matched corrosions will contribute to pipelines risk analysis and integrity management. On

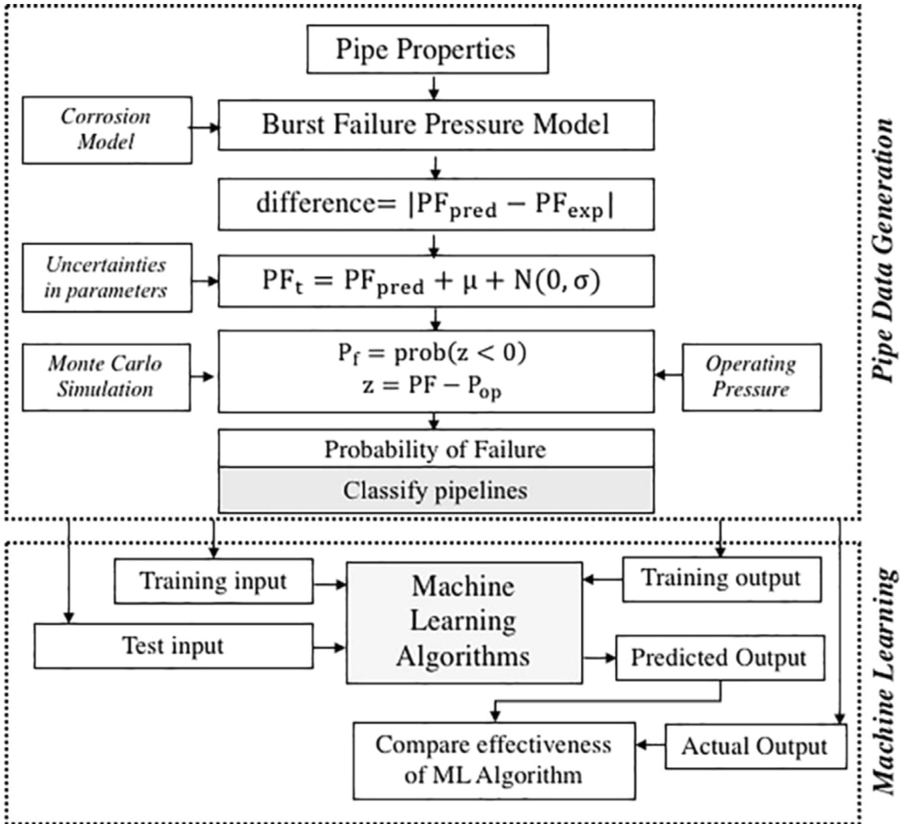

Fig. 14. The framework proposed by Mazumder et al. for failure risk analysis (Mazumder et al., 2021).

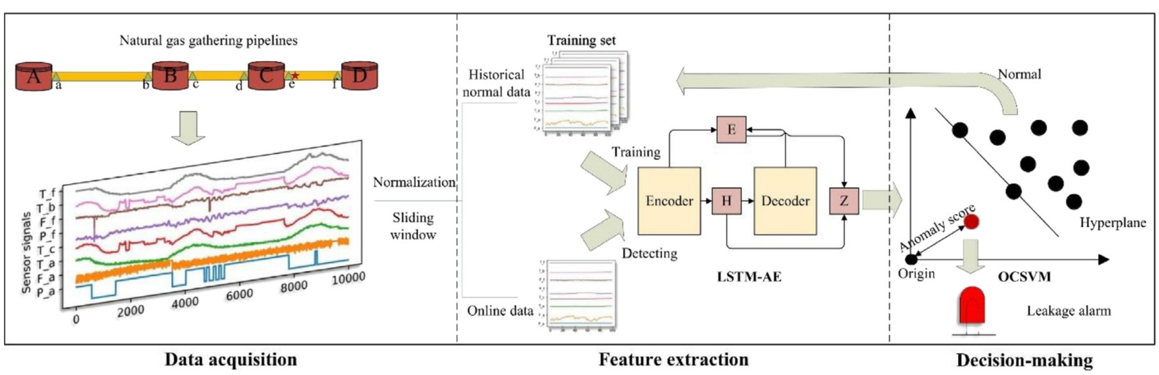

Fig. 15. The framework of the proposed method by Zuo et al. for detecting leaks (Zuo et al., 2022).

the other hand, real-time leak detection in oil and gas pipelines is vital for industry safety. Traditional model-based techniques are impractical, leading to increased interest in data-driven methods. However, most of the existing approaches require labeled data, which is challenging to obtain for rare leaks. To address this, Zuo et al. (2022) proposed a semi-supervised method using an improved long short-term memory autoencoder (LSTM-AE) network and a one-class support vector machine (OCSVM) as illustrated in Fig. 15. The LSTM-AE learns features from normal parameter datasets, and the OCSVM calculates a score to detect leaks. Evaluation on real gas pipelines showed 98% accuracy and 99% area under the curve, confirming the effectiveness of the developed approach.

A new method called Sparrow Search Algorithm and Convolutional Neural Network (SSA-CNN) for detecting oil pipeline leaks was proposed by Li et al. (2022a). The SSA-CNN method initially converts the input time series data into a two-dimensional matrix and then compares the classification conditions using different convolution kernel sizes and pooling sizes. The SSA algorithm was employed to optimize the CNN parameters. It was found that using two-dimensional data as input improved the neural network's ability to extract features compared to traditional ML methods. The proposed SSA-CNN method achieved an accuracy rate of 98.67% which is higher than that of traditional ML methods and further improves the classification capability of CNN. Spandonidis et al. (2022) conducted a study using a semi-supervised deep-learning approach to detect oil leaks in pipelines. Two methodologies were implemented: a 2D-CNN for supervised classification of spectrograms, and LSTM AE for unsupervised leakage detection. The combined model outperformed the benchmark and individual components, demonstrating its effectiveness for pipeline leak detection. Table 3 presents the summary of some of the studies that combined more than a machine learning technique to improve performance and prediction accuracy. Similar to studies that used ANN, most of the existing studies are on corrosion and still there is a knowledge gap in the applications of hybrid machine learning models for predicting other most common failures of oil and gas pipelines. Overall and based on the aforementioned studies, it can be said that hybrid ML techniques showed a significant accuracy in the prediction of oil and gas pipeline failures compared to the separate models.

Based on the relevant literature have been reviewed in Section 3.3.2, it can be said that understanding the strengths, weaknesses, and applicability of ANNs, SVMs, and hybrid models is crucial for predicting defects in oil and gas pipelines. They provide distinct advantages and limitations, making them valuable tools for precise oil and gas pipeline defect prediction and prevention to some extent. ANNs are advantageous for predicting defects in oil and gas pipelines due to their ability to learn complex patterns, handle large amounts of data, and make predictions based on incomplete or noisy data. However, they have limitations, including being "black-box" models that are difficult to interpret, requiring substantial training data, time-consuming to train and optimize, being computationally intensive, and being prone to overfitting if not properly regularized. Similarly, SVMs are useful for predicting defects in oil and gas pipelines as they handle high-dimensional data, non-linear relationships, and provide a clear margin of separation between classes. They are also robust to outliers. However, SVMs can be sensitive to hyperparameters and the choice of kernel functions. They can be computationally expensive with large datasets and may struggle with overlapping classes or imbalanced class distributions. Hybrid machine learning models, which combine different algorithms such as ANNs and SVMs, offer strengths that can be beneficial for predicting defects in oil and gas pipelines. Hybrid models can leverage the strengths of individual algorithms, enhancing overall prediction accuracy. They can handle complex patterns and relationships and mitigate the weaknesses of individual algorithms. However, developing and implementing hybrid models can be a challenging and time-consuming process. The performance of hybrid models heavily depends on the selection and integration of individual algorithms.

In summary, ANNs can predict future defects based on historical data, but their performance relies on data quality and quantity. SVMs handle high-dimensional data and non-linear relationships but need careful tuning and balanced datasets. Hybrid models enhance prediction accuracy by combining algorithms, but their design and implementation require careful consideration and evaluation.

4 Discussions

According to the studies reviewed in this SLR, the following subsections provide three aspects of literature including the motivations behind applying ML techniques in the development of predictive models for oil and gas pipeline failures, the challenges encountered in using ML techniques successfully in oil and gas pipeline defects prediction and recommendations and future directions to overcome and mitigate such difficulties.

4.1 Motivations

One of the motivations for using ML for detecting the failures of oil and gas pipelines is the challenges faced by oil and gas industries to inspect oil and gas pipelines that transport hydrocarbons over a very long distance using the most common non-destructive evaluations techniques based on ultrasonic testing, magnetic flux leakages, etc., which are techniques that required stopping the current operations of the pipelines resulting in a reduction of the financial profit (Bastian et al., 2019; De Masi et al., 2015). Besides, the manual inspection method is one of the conventional alternatives for failure detection; however, it is time-consuming and not efficient and practical in case of hazardous locations. In this method closed circuit TV (CCTV) cameras are installed to capture the failure images and then expert humans conduct a manual checking which is considered very tiresome and time-consuming and also not very accurate (Bastian et al., 2019). Thus, there is a need for a computer vision-based system that can be used to capture images from dynamic moving camera systems and automatically detect failures.

Furthermore, the applications of hybrid machine learning models by combining more ML models showed a significant improvement in the accuracies and efficiencies of the developed models for predicting the oil and gas pipeline defects compared to the standalone model used. Such HML techniques are widely applied for predicting pipeline corrosion, however, it is still considered a research gap for improvement of most other pipeline failure defects such as third-party, natural defects, mechanical defects and so on. Therefore, encouraged findings from the application of HML for corrosion prediction could be a motivation and a solid background guiding the extent of application of such technique for prediction of the rest of pipeline failures.

4.2 Challenges

One of the main challenges of oil and gas pipeline failures predictions is inadequate understanding and incorrect diagnosis of the problem of the exact failure through the studying of the nature of the oil and gas materials, from the design and materials aspect, implementation aspect, chemistry aspect, environmental aspect and flow mechanics aspect. Then based on the comprehensive investigation and specifying the exact issues, the role of the collection of accurate data and applying ML come to develop prediction models that can reflect the real situation. Another main challenge is dealing with, processing and analyzing the datasets from different resources. Most of such challenges are due to instrument performance, adopted technology for data collection and changing the reporting criteria. Besides, the earlier data processing through hardware or software can be another reason for such challenges (Khan et al., 2021; Soomro et al., 2022b). To contribute to solving some of the aforementioned challenges, a holistic approach in terms of a framework that includes data cleaning and statistical analysis has been developed to overcome the challenges of accurate and credible interpretation of large databases prior to being used for developing predictive models (Khan et al., 2021).

Furthermore, to evaluate the pipeline failures, another challenge is the availability of data due to the complication of the oil and gas pipeline operation process as a result of the fluctuation of operating conditions and different product components and properties. Besides, the existing data is very limited. Therefore, there is a need for probabilistic models to address the lack of information in this regard (Liu et al., 2021). Moreover, dozens of ML algorithms are available and each of them follows a different theory of learning. It cannot be nominated as one of those algorithms as the best among the others or even there is no one algorithm that can fit all (Yin et al., 2021). That means choosing the right algorithm is very critical and considered another challenge for applications of ML in oil and gas pipeline failure predictions.

However, extensive studies in literature used the experimental, simulations and field inspection as the main resource for data that needs to develop ML models for oil and gas pipelines failures prediction, there is a need to further extend the possibility of using different simulation techniques such as finite element method (FEM) at various stress, strain, environmental, operation conditions and pipeline characteristics to further minimize the

Table 3
Summary of studies on prediction of oil and gas pipeline failures using hybrid ML techniques.

Reference	ML technique used	Input parameters	Output evaluated	Remarks
Liu et al. (2021)	– Classification tree techniques – ANN (Feed forward) with two hidden layers – Bayesian Network to check uncertainty	Pipe diameter, radius of curvature, steel of hardness, particle diameter, particle hardness, particle density, fluid density, fluid viscosity and fluid velocity	Predict the erosion rates	– Dataset used was obtained from a Monte-Carlo method simulation. – Classification tree showed to be the faster technique used for erosion rate detection, whereas ANN provides a comprehensive sensitivity evaluation. – ANN back propagation with two hidden layers, 10 neurons, showed to be the optimal structure for pipeline leakage very well.
Yin et al. (2021)	ML (Multilayer perceptron, SVM and random forest) and Fuzzy	Safety effects, transportation interruption influence, equipment maintenance and environmental effects	Failure criticality of pipelines	– It was found that the random forest model showed the best prediction capability compared to other ML models used. Besides, the effects of safety and environmental factors exhibited the biggest impact on failure criticality.
Mazzella et al. (2019)	Hybrid ML techniques (ANN, extreme boosted trees and generalized linear model)	Focused on the environmental factors (chloride pollution, annual average temperature, sulfide pollution, time of wetness and number of years below than zero degree), besides the pipe manufacturer, actual diameter and year of mill run.	Underground oil and gas pipeline's corrosion rate	– Data was obtained from North American pipeline operator – All three models showed a significant prediction for the corrosion rate of pipeline with a high degree of accuracy
Li et al. (2022a)	Sparrow Search Algorithm and Convolutional Neural Network (SSA-CNN)	The size of the convolutional layer's kernel, the number of neurons, the rate of learning, and the number of iterations of the CNN.	Pipeline leaks	– Two-dimensional input data improved feature extraction in neural networks compared to traditional ML methods – The proposed SSA-CNN method achieved a higher accuracy rate of 98.67% and enhanced the classification capability of CNN.
Priyanka et al. (2021)	ML (manifold learning method, kernel based SVM algorithm)	At wide pressure range	Predict the risk probability rate	– A new technique was proposed to handle the oil and gas pipeline failure estimation with respect to pressure factors. As claimed, this technique is consistent with the most recently developed IoT technologies to be applied in the real-world oil and gas pipeline.
Liao et al. (2012)	ANN, ANN-PSO and ANN-GA	Liquid hold up, maximum wall stresses, heat transfer coefficients of pipeline wall, superficial velocities, maximum shear stress in the wall, deposition rate and pipe angle	Internal corrosion rates of gas pipeline	– Data obtained from the field inspection – ANN-PSO showed the best accuracy among the three developed models
Spandonidis et al. (2022)	Semi-supervised deep-learning (a 2D-CNN for supervised classification of spectrograms, and LSTM AE for unsupervised)	The sensor spacing, leakage-sensor separation, and leakage diameter.	Leakage detection	– The combined model surpassed both the benchmark and individual components, indicating its efficacy for detecting pipeline leaks.
Ossai (2019)	– HML (Feed forward ANN with particle swarm optimization (PSO), Random Forest (RF), Gradient Boosting Machine (GBN), and Deep Neural Network (DNN)) – Principle Component Analysis (PCA)	Operating pressure, temperature, CO_2 pressure, sulphate ion concentrations, chloride ion concentrations, oil production rates, gas production rates, alkalinity concentration, iron content and calcium concentration	Corrosion defect depth of aged pipelines	– Field measurements have been conducted to collect the data – Results showed that the accuracy of the prediction from the combined ML modeling with PCA was found to be 3.52 to 5.32 times of that using only PCA. – GBN with PCA exhibited the best accuracy among all other HML
Aslam (2018)	ANN, Genetic Algorithm (GA) and Fuzzy Logic (FL)	External parameters (external temperature, weather patterns and the elevation) and internal stress parameters (gas composition, hydrocarbon composition, velocity and flow rate)	Leak and corrosion prediction	– Data collected from the field measurements – Developed HML model showed to be practical for accurate prediction of leak and corrosion of pipelines

(*continued on next page*)

Table 3 (*continued*).

Reference	ML technique used	Input parameters	Output evaluated	Remarks
Ossai (2020)	Feed-forward ANN using sub cluster neural network (SCNN) and PSO	Operating pressure, temperature, basic sediments and water, CO_2 partial pressure, gas specific gravity and oil and gas prediction rates	Corrosion defect depth and burst pressure	– The developed model exhibited a degree of accuracy with a predicted correlation coefficient of 92%
Sinha and Pandey (2002)	Probabilistic neural network (PNN) and back propagation neural network (BPNN)	Pipe wall thickness, yield strength of pipe, cracks depth, pipe diameter, the standard deviation of crack depth and operation pressures	Probability of failure (POF) of oil and gas pipeline	– Field observation data were collected using a magnetic flux leakage tool – Proposed model trained by PNN showed better accuracy compared to BPNN based model.
Liu et al. (2012)	PSO-SVM	Acid number of oils, sulfur in crude oil, temperatures, pressures and velocities	Internal corrosion rate of oil pipelines	– The proposed model has been compared to BPNN and found that PSO-SVM resulted in better accuracy with a maximum error of 0.6%
Zuo et al. (2022)	Long short-term memory autoencoder (LSTM-AE) network and a one-class support vector machine (OCSVM)	Pressure, flow and temperature	Real-time leak detection	– The developed approach demonstrated its effectiveness with an accuracy of 98% and an area under the curve of 99%.
Peng et al. (2021)	Support vector regression (SVR), chaos particle swarm optimization (CPSO) and principle component analysis (PCA)	pH, flow rate, temperature, partials CO_2 pressures, pressure, wall shaving stress and liquid hold	Corrosion rates for multiphase flow pipelines	– Data collected from the 5.5 km submarine oil and gas pipeline in Hainan, China – SVR-CPSO-PCA showed a better degree of prediction and accuracy compared to the SVR-based model with an error of 0.0083.
Phan and Duong (2021)	Adaptive neuro fuzzy inference system (ANFIS) with and without PCA	Wall thickness, pipe diameter, corrosion defect length and corrosion defect depth,	Burst pressure of pipeline	– The results showed that models with PCA exhibited better accuracy compared to that without PCA

need for the field and experimental inspections which is definitely required more time and cost. Meanwhile, it is important to consider whether the data obtained through field and experimental methods is more representative compared to simulation and modeling. models developed based on the data obtained from the simulation need to be validated to ensure the behavior simulated is very close to the real pipeline failure behavior. It is also essential to carefully consider the representativeness and reliability of the data used for training and validation. Balancing the trade-off between the costs and benefits of different data collection methods, such as field inspections, experimental testing, and simulation, will be crucial to ensure the accuracy and generalizability of the deep learning models. Besides, in-depth studies including the micro and molecular level on the different oil and gas pipeline defects to predict early failure and can provide adequate solutions earlier is another challenge and research gaps should be studied in the future extensively.

4.3 Recommendations and future directions

However, many studies applied ML techniques to predict the corrosion defects of oil and gas pipelines, but very limited studies addressed other failures such as weldment defects with their different types (weldment materials failure, porosity, overlayers, weldment cracking and etc.), Pipelines materials properties and manufacturing defects, environmental condition defects, stress and strain (including tensile, shear and bending) based defects, pressure or burst pressure defects. Therefore, applications of ML techniques still need to be further extensively used to come up with accurate systems that can be used for oil and gas pipelines failures predictions in order to improve the integrity of oil and gas pipelines and enhance the control of overall health monitoring system of oil and gas industries. Combining more than the approach of ML such as ANN, SVM and other algorithms could be another trend for future studies to come up with higher accuracy.

Based on the recommendations reported in the literature, lab datasets should be extended either experimentally or use various available simulators to improve the predictive capability of existing models. Furthermore, the hybrid ML technique should be adopted to solve the existing issues such as pressure fluctuation in the oil and gas pipeline which has a significant effect on the operations. That also can be done by incorporating the ML algorithms into the mechanistic models (Kanin et al., 2019). It was also recommended that, in order to develop an accurate prediction and detection of pipeline failure models, hybrid techniques of software-based techniques and hardware-based techniques should be adopted to come up with low errors and adequate accuracy. So the combination of different techniques and technologies is recommended to improve the reliability and accuracy of pipeline failure detection (Qin et al., 2023; Vandrangi et al., 2022). Although developed predictive models for the criticality index of oil and gas pipelines by Yin et al. (2021) can be useful for researchers and industries that are interested in applications of ML for modeling oil and gas pipelines failures, future research was recommended on multi-level evaluation indicators in deeper with respect to each factor at different situations to come up with more practical models. There is also a need for comprehensive research on the applications of ML techniques for detecting the interactive defects of oil and gas pipelines based on the timing and locations of these defects (Liu and Bao, 2022).

5 Conclusions

This systematic review paper focuses on the use of machine and deep learning techniques, specifically artificial neural networks (ANNs), support vector machines (SVMs) and hybrid machine learning (HML) algorithms, for predicting various pipeline failures in the oil and gas industry. Unlike previous reviews, it specifically examines how these techniques are applied, including the parameters and data reliability involved. The paper provides a comprehensive discussion of the motivations and challenges of using ML for predicting different types of pipeline defects. It also includes a bibliometric analysis that highlights commonly used ML techniques, the types of failures investigated, and the experimental tests conducted. Detailed information, summarized in tables, is provided on different failure types, commonly used ML algorithms, and available data resources, along with critical discussions. According to the results and analysis from this systematic review, the following conclusions can be drawn:

- The effectiveness of machine learning methods, particularly accuracy, is influenced by numerous factors, including variations in dataset sizes, data sources, pipeline types, data types, pre-processing algorithms, and machine learning models.
- In comparison to conventional techniques, machine learning methods such as ANN, SVM, and HML demonstrated low complexity and achieved high to very high accuracy in detecting pipeline defects. However, it is important to note that a substantial amount of data is necessary for optimal performance.
- The majority of studies found in the literature focused on addressing corrosion defects in oil and gas pipelines. However, there were only a few studies that explored the application of machine learning on other types of failures, such as third-party, mechanical, natural, and other failures.
- Through the bibliometric analysis, it was observed that the most commonly used machine learning techniques in the literature for predicting pipeline defects are ANN and SVM. However, there is a scarcity of research that has explored the utilization of other machine-learning techniques, such as CNN, and KNN.
- In generating laboratory data for pipeline defect detection, nondestructive examination and magnetic flux leakage tests are the most commonly utilized methods. However, it is suggested that incorporating more advanced simulations like finite element methods can enhance the accuracy of machine learning models by providing more up-to-date and reliable data.
- The parameters most frequently employed in the literature for predictive machine learning models were operating pressure, pH, temperature, CO_2 pressure, velocity, pipe wall thickness, defect depth and length, diameter, flow rate, density of liquid, and liquid hold-up.
- ML models developed using ANN and/or SVM techniques exhibited a strong correlation with field, laboratory, and simulation-based data. However, to enhance the accuracy of the models and enable the detection of various types of failures, it is recommended to explore the application of hybrid ML techniques that have not been extensively studied in the existing literature.
- Machine learning in the oil and gas pipeline sector is advantageous for its ability to learn from diverse data sources and create accurate models. However, the challenge lies in generating sufficient clean data for training these models.
- Conducting experiments with large datasets can enhance the generalization capability of defects detection methods, enabling a more comprehensive analysis of pipeline defects.
- There are many different ML algorithms, each with its own theory of learning. However, there is no one-size-fits-all algorithm that is considered the best. This makes it challenging to choose the right algorithm for predicting oil and gas pipeline failures. To overcome this challenge, future research should prioritize the optimization of ML methods and the enhancement of accuracy.
- While optimization algorithms improve detection accuracy, they also increase training time, making it challenging to continuously update the model for real-time detection. Future research should focus on reducing model training time and enabling continuous updates while ensuring real-time detection.
- More research is needed to establish a sustainable machine-learning system that can effectively monitor all oil and gas pipeline failures, considering the environment. This system should automate processes from data collection to decision-making for repairing identified defects.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The authors wish to acknowledge Yayasan Universiti Teknologi Petronas, Malaysia for supporting this work through the research grants (015LC0-286 and 015LC0-308).

References

Adegboye, M.A., Fung, W.-K., Karnik, A., 2019. Recent advances in pipeline monitoring and oil leakage detection technologies: Principles and approaches. Sensors 19 (11), 2548.
Afebu, K., Abbas, A., Nasr, G., Kadir, A., 2015. Integrated leak detection in gas pipelines using OLGA simulator and artificial neural networks. In: Abu Dhabi International Petroleum Exhibition and Conference. OnePetro.
Agency, C.I., 2020. The World Factbook 2020. Central Intelligence Agency.
Al-Amin, M., Zhou, W., Zhang, S., Kariyawasam, S., Wang, H., 2012. Bayesian model for calibration of ILI tools. In: International Pipeline Conference. American Society of Mechanical Engineers, pp. 201–208.
Alamri, A.H., 2022. Application of machine learning to stress corrosion cracking risk assessment. Egypt. J. Pet. 31 (4), 11–21.
Amini, M., Sharifani, K., Rahmani, A., 2023. Machine learning model towards evaluating data gathering methods in manufacturing and mechanical engineering. Int. J. Appl. Sci. Eng. Res. 15 (2023), 349–362.
Aslam, N., 2018. Artificial intelligence based algorithm for predicting pipeline leak and corrosion detection. Google Patents.
Bastian, B.T., Jaspreeth, N., Ranjith, S.K., Jiji, C., 2019. Visual inspection and characterization of external corrosion in pipelines using deep neural network. NDT E Int. 107, 102134.
Behnood, A., Daneshvar, D., 2020. A machine learning study of the dynamic modulus of asphalt concretes: An application of M5P model tree algorithm. Constr. Build. Mater. 262, 120544.
Biezma, M., Andrés, M., Agudo, D., Briz, E., 2020. Most fatal oil & gas pipeline accidents through history: A lessons learned approach. Eng. Fail. Anal. 110, 104446.
Briner, R.B., Denyer, D., 2012. Systematic review and evidence synthesis as a practice and scholarship tool.
Carvalho, A., Rebello, J., Sagrilo, L., Camerini, C., Miranda, I., 2006. MFL signals and artificial neural networks applied to detection and classification of pipe weld defects. Ndt E Int. 39 (8), 661–667.
Chamkalani, A., et al., 2013. Soft computing method for prediction of CO2 corrosion in flow lines based on neural network approach. Chem. Eng. Commun. 200 (6), 731–747.
Chen, H., et al., 2022. Application of machine learning to evaluating and remediating models for energy and environmental engineering. Appl. Energy 320, 119286.
Chin, K.T., Arumugam, T., Karuppanan, S., Ovinis, M., 2020. Failure pressure prediction of pipeline with single corrosion defect using artificial neural network. Pipeline Sci. Technol. 4 (1), 3.
Crawley, F., 2020. In: Crawley, F. (Ed.), 12—Failure Modes and Effects Analysis (FMEA) and Failure Modes, Effects and Criticality Analysis (FMECA). a Guide To Hazard Identification Methods, second ed. pp. 103–109.
da Cruz, R.P., da Silva, F.V., Fileti, A.M.F., 2020. Machine learning and acoustic method applied to leak detection and location in low-pressure gas pipelines. Clean Technol. Environ. Policy 22, 627–638.
Dai, L., Wang, D., Wang, T., Feng, Q., Yang, X., 2017. Analysis and comparison of long-distance pipeline failures. J. Pet. Eng. (2017).
De Masi, G., Gentile, M., Vichi, R., Bruschi, R., Gabetta, G., 2015. Machine learning approach to corrosion assessment in subsea pipelines. In: OCEANS 2015-Genova. IEEE, pp. 1–6.

De Masi, G., Vichi, R., Gentile, M., Bruschi, R., Gabetta, G., 2014. A neural network predictive model of pipeline internal corrosion profile. In: Proceedings of the 1st International Conference on Systems Informatics, Modeling and Simulation. pp. 18–23.
Din, M.M., et al., 2015. An artificial neural network modeling for pipeline corrosion growth prediction. ARPN J. Eng. Appl. Sci. 10 (2), 512–519.
Du, J., et al., 2023. Deeppipe: Theory-guided prediction method based automatic machine learning for maximum pitting corrosion depth of oil and gas pipeline. Chem. Eng. Sci. 118927.
Eastvedt, D., Naterer, G., Duan, X., 2022. Detection of faults in subsea pipelines by flow monitoring with regression supervised machine learning. Process Saf. Environ. Prot. 161, 409–420.
Fang, Y., Rasel, M., Richmond, P.C., 2020. Consequence risk analysis using operating procedure event trees and dynamic simulation. J. Loss Prev. Process Ind. 67, 104235.
Ferreira, A.D.M., Afonso, S.M., Willmersdorf, R.B., Lyra, P.R., 2021. Multiresolution analysis and deep learning for corroded pipeline failure assessment. Adv. Eng. Softw. 162, 103066.
Girgin, S., Krausmann, E., 2016. Historical analysis of US onshore hazardous liquid pipeline accidents triggered by natural hazards. J. Loss Prev. Process Ind. 40, 578–590.
Hegde, J., Rokseth, B., 2020. Applications of machine learning methods for engineering risk assessment–A review. Saf. Sci. 122, 104492.
Ho, M., El-Borgi, S., Patil, D., Song, G., 2020. Inspection and monitoring systems subsea pipelines: A review paper. Struct. Health Monit. 19 (2), 606–645.
Hu, X., Barker, R., Neville, A., Gnanavelu, A., 2011. Case study on erosion–corrosion degradation of pipework located on an offshore oil and gas facility. Wear 271 (9–10), 1295–1301.
Jiang, F., Dong, S., 2020. Collision failure risk analysis of falling object on subsea pipelines based on machine learning scheme. Eng. Fail. Anal. 114, 104601.
Kanin, E., Osiptsov, A., Vainshtein, A., Burnaev, E., 2019. A predictive model for steady-state multiphase pipe flow: Machine learning on lab data. J. Pet. Sci. Eng. 180, 727–746.
Karimi, S., Shirazi, S.A., McLaury, B.S., 2017. Predicting fine particle erosion utilizing computational fluid dynamics. Wear 376, 1130–1137.
Karpatne, A., et al., 2017. Theory-guided data science: A new paradigm for scientific discovery from data. IEEE Trans. Knowl. Data Eng. 29 (10), 2318–2331.
Khan, F., Yarveisy, R., Abbassi, R., 2021. Cross-country pipeline inspection data analysis and testing of probabilistic degradation models. J. Pipeline Sci. Eng. 1 (3), 308–320.
Koch, G., et al., 2016. International Measures of Prevention, Application, and Economics of Corrosion Technologies Study, Vol. 216. NACE international, pp. 2–3.
Layouni, M., Hamdi, M.S., Tahar, S., 2017. Detection and sizing of metal-loss defects in oil and gas pipelines using pattern-adapted wavelets and machine learning. Appl. Soft Comput. 52, 247–261.
Layouni, M., Tahar, S., Hamdi, M.S., 2014. A survey on the application of neural networks in the safety assessment of oil and gas pipelines. In: 2014 IEEE Symposium on Computational Intelligence for Engineering Solutions (CIES). IEEE, pp. 95–102.
Li, Q., Shi, Y., Lin, R., Qiao, W., Ba, W., 2022a. A novel oil pipeline leakage detection method based on the sparrow search algorithm and CNN. Measurement 204, 112122.
Li, X., Wang, J., Chen, G., 2022b. A machine learning methodology for probabilistic risk assessment of process operations: A case of subsea gas pipeline leak accidents. Process Saf. Environ. Prot. 165, 959–968.
Liao, K., Yao, Q., Wu, X., Jia, W., 2012. A numerical corrosion rate prediction method for direct assessment of wet gas gathering pipelines internal corrosion. Energies 5 (10), 3892–3907.
Liao, Q., et al., 2022. Innovations of carbon-neutral petroleum pipeline: A review. Energy Rep. 8, 13114–13128.
Little, R.J., Rubin, D.B., 2019. Statistical Analysis with Missing Data, Vol. 793. John Wiley & Sons.
Liu, G., Ayello, F., Vera, J., Eckert, R., Bhat, P., 2021. An exploration on the machine learning approaches to determine the erosion rates for liquid hydrocarbon transmission pipelines towards safer and cleaner transportations. J. Clean. Prod. 295, 126478.
Liu, Y., Bao, Y., 2022. Review on automated condition assessment of pipelines with machine learning. Adv. Eng. Inform. 53, 101687.
Liu, W., Chen, Z., Hu, Y., 2022. XGBoost algorithm-based prediction of safety assessment for pipelines. Int. J. Press. Vessels Pip. 197, 104655.
Liu, E., Guo, B., Wang, M., Ma, X., Peng, Y., 2020. Analysis of the water-filling process for crude oil pipelines with a large drop in height. Energy Sci. Eng. 8 (6), 2100–2115.
Liu, H., Liu, Z., Taylor, B., Dong, H., 2019b. Matching pipeline in-line inspection data for corrosion characterization. NDT E Int. 101, 44–52.
Liu, E., Lv, L., Yi, Y., Xie, P., 2019a. Research on the steady operation optimization model of natural gas pipeline considering the combined operation of air coolers and compressors. IEEE Access 7, 83251–83265.
Liu, J., Wang, H., Yuan, Z., 2012. Forecast model for inner corrosion rate of oil pipeline based on PSO-SVM. Int. J. Simul. Process Model. 7 (1–2), 74–80.
Liu, R., Xiong, H., Wu, X., Yan, S., 2014. Numerical studies on global buckling of subsea pipelines. Ocean Eng. 78, 62–72.
Liu, P., et al., 2023. A CNN-based transfer learning method for leakage detection of pipeline under multiple working conditions with AE signals. Process Saf. Environ. Prot. 170, 1161–1172.
Loebbecke, C., Picot, A., 2015. Reflections on societal and business model transformation arising from digitization and big data analytics: A research agenda. J. Strateg. Inf. Syst. 24 (3), 149–157.
Lu, H., Iseley, T., Matthews, J., Liao, W., Azimi, M., 2021. An ensemble model based on relevance vector machine and multi-objective salp swarm algorithm for predicting burst pressure of corroded pipelines. J. Pet. Sci. Eng. 203, 108585.
Luo, Z., Hu, X., Gao, Y., 2013. Corrosion research of wet natural gathering and transportation pipeline based on SVM. In: ICPTT 2013. Trenchless Technology, pp. 964–972.
Ma, H., et al., 2023. A new hybrid approach model for predicting burst pressure of corroded pipelines of gas and oil. Eng. Fail. Anal. 149, 107248.
Mazumder, R.K., Salman, A.M., Li, Y., 2021. Failure risk analysis of pipelines using data-driven machine learning algorithms. Struct. Saf. 89, 102047.
Mazzella, J., Hayden, T., Krissa, L., Tsaprailis, H., 2019. Estimating corrosion growth rate for underground pipeline: a machine learning based approach. In: CORROSION 2019. OnePetro.
Mohamed, A., Hamdi, M.S., Tahar, S., 2015. A machine learning approach for big data in oil and gas pipelines. In: 2015 3rd International Conference on Future Internet of Things and Cloud. IEEE, pp. 585–590.
Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., Group*, P., 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann. Intern. Med. 151 (4), 264–269.
Mohtadi-Bonab, M., 2019. Effects of different parameters on initiation and propagation of stress corrosion cracks in pipeline steels: a review. Metals 9 (5), 590.
Murphy, K.P., 2012. Machine Learning: A Probabilistic Perspective. MIT Press.
Nayak, N., Anarghya, A., Adhoubi, M.Al., 2020. A study on the behavior of CO2 corrosion on pipeline using computational fluid dynamics, experimental and artificial neural network approach. Eng. Res. Express 2 (2), 025012.
Ossai, C.I., 2019. A data-driven machine learning approach for corrosion risk assessment—a comparative study. Big Data Cogn. Comput. 3 (2), 28.
Ossai, C.I., 2020. Corrosion defect modelling of aged pipelines with a feed-forward multi-layer neural network for leak and burst failure estimation. Eng. Fail. Anal. 110, 104397.
Peng, S., Zhang, Z., Liu, E., Liu, W., Qiao, W., 2021. A new hybrid algorithm model for prediction of internal corrosion rate of multiphase pipeline. J. Natl. Gas Sci. Eng. 85, 103716.
Phan, H.C., Duong, H.T., 2021. Predicting burst pressure of defected pipeline with principal component analysis and adaptive neuro fuzzy inference system. Int. J. Press. Vessels Pip. 189, 104274.
Piao, G., Guo, J., Hu, T., Leung, H., Deng, Y., 2019. Fast reconstruction of 3-D defect profile from MFL signals using key physics-based parameters and SVM. NDT E Int. 103, 26–38.
Priyanka, E., Thangavel, S., Gao, X.-Z., Sivakumar, N., 2021. Digital twin for oil pipeline risk estimation using prognostic and machine learning techniques. J. Ind. Inf. Integr. 100272.
Qin, G., et al., 2023. A hybrid machine learning model for predicting crater width formed by explosions of natural gas pipelines. J. Loss Prev. Process Ind. 82, 104994.
Rachman, A., Zhang, T., Ratnayake, R.C., 2021. Applications of machine learning in pipeline integrity management: A state-of-the-art review. Int. J. Press. Vessels Pip. 193, 104471.
Ramadhan, R.A., Heatubun, Y.R., Tan, S.F., Lee, H.-J., 2021. Comparison of physical and machine learning models for estimating solar irradiance and photovoltaic power. Renew. Energy 178, 1006–1019.
Ren, C.-y., Qiao, W., Tian, X., 2012. Natural gas pipeline corrosion rate prediction model based on BP neural network. In: Fuzzy Engineering and Operations Research. Springer, pp. 449–455.
Roth, P.L., 1994. Missing data: A conceptual review for applied psychologists. Pers. Psychol. 47 (3), 537–560.
Rother, E.T., 2007. Systematic literature review x narrative review. Acta Paul. Enferm. 20, v–vi.
Santoso, B., Indarto, Deendarlianto, ., 2014. Pipeline leak detection in two phase flow based on fluctuation pressure difference and artificial neural network (ANN). Appl. Mech. Mater. 493, 186–191.
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. J. Natl. Gas Sci. Eng. 99, 104425.
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.
Shaik, N.B., et al., 2022. Recurrent neural network-based model for estimating the life condition of a dry gas pipeline. Process Saf. Environ. Prot. 164, 639–650.

Silva, R., Guerreiro, J., Loula, A., 2007. A study of pipe interacting corrosion defects using the FEM and neural networks. Adv. Eng. Softw. 38 (11–12), 868–875.
Sinha, S.K., Pandey, M.D., 2002. Probabilistic neural network for reliability assessment of oil and gas pipelines. Comput.-Aided Civ. Infrastruct. Eng. 17 (5), 320–329.
Soomro, A.A., et al., 2022a. Integrity assessment of corroded oil and gas pipelines using machine learning: A systematic review. Eng. Fail. Anal. 131, 105810.
Soomro, A.A., et al., 2022b. A review on Bayesian modeling approach to quantify failure risk assessment of oil and gas pipelines due to corrosion. Int. J. Press. Vessels Pip. 104841.
Spandonidis, C., Theodoropoulos, P., Giannopoulos, F., 2022. A combined semi-supervised deep learning method for oil leak detection in pipelines using IIoT at the edge. Sensors 22 (11), 4105.
Su, Y., Li, J., Yu, B., Zhao, Y., Yao, J., 2021. Fast and accurate prediction of failure pressure of oil and gas defective pipelines using the deep learning model. Reliab. Eng. Syst. Saf. 216, 108016.
Sukarno, P., et al., 2007. Leak detection modeling and simulation for oil pipeline with artificial intelligence method. J. Eng. Technol. Sci. 39 (1), 1–19.
Sun, T., et al., 2022. Magnetic anomaly detection of adjacent parallel pipelines using deep learning neural networks. Comput. Geosci. 159, 104987.
Vadyala, S.R., Betgeri, S.N., Matthews, J.C., Matthews, E., 2022. A review of physics-based machine learning in civil engineering. Results Eng. 13, 100316.
Valentin de Oliveira, T., 2018. Leakage Prevention and Detection in Pipelines Utilizing a Wireless Information and Communication Network. Schulich School of Engineering.
Valizadeh, S., Moshiri, B., Salahshoor, K., 2009. Multiphase pipeline leak detection based on fuzzy classification. In: AIP Conference Proceedings. American Institute of Physics, pp. 72–80.
Vandrangi, S.K., Lemma, T.A., Mujtaba, S.M., Ofei, T.N., 2022. Developments of leak detection, diagnostics, and prediction algorithms in multiphase flows. Chem. Eng. Sci. 248, 117205.
Wang, M., Cheng, J.C., 2020. A unified convolutional neural network integrated with conditional random field for pipe defect segmentation. Comput.-Aided Civ. Infrastruct. Eng. 35 (2), 162–177.
Wang, H., Yajima, A., Liang, R.Y., Castaneda, H., 2015. A Bayesian model framework for calibrating ultrasonic in-line inspection data and estimating actual external corrosion depth in buried pipeline utilizing a clustering technique. Struct. Saf. 54, 19–31.
Wang, N., Zhang, D., Chang, H., Li, H., 2020. Deep learning of subsurface flow via theory-guided neural network. J. Hydrol. 584, 124700.
Wasim, M., Djukic, M.B., 2022. External corrosion of oil and gas pipelines: A review of failure mechanisms and predictive preventions. J. Natl. Gas Sci. Eng. 100, 104467.
Wei, B., Xu, J., Sun, C., Cheng, Y.F., 2022. Internal microbiologically influenced corrosion of natural gas pipelines: A critical review. J. Natl. Gas Sci. Eng. 102, 104581.
Willersrud, A., Blanke, M., Imsland, L., 2015. Incident detection and isolation in drilling using analytical redundancy relations. Control Eng. Pract. 41, 1–12.
Witte, R., Witte, J., 2017. Statistics, Vol. 501. John Wiley & Sons. Inc, Hoboken.
Worrell, C., Luangkesorn, L., Haight, J., Congedo, T., 2019. Machine learning of fire hazard model simulations for use in probabilistic safety assessments at nuclear power plants. Reliab. Eng. Syst. Saf. 183, 128–142.
Wright, R.W., Brand, R.A., Dunn, W., Spindler, K.P., 2007. How to write a systematic review. Clin. Orthop. Rel. Res. (1976-2007) 455, 23–29.
Wu, T., et al., 2023. Oil pipeline leakage monitoring developments in China. J. Pipeline Sci. Eng. 100129.
Xiao, R., Hu, Q., Li, J., 2019. Leak detection of gas pipelines using acoustic signals based on wavelet transform and support vector machine. Measurement 146, 479–489.
Xiao, Y., Watson, M., 2019. Guidance on conducting a systematic literature review. J. Plan. Educ. Res. 39 (1), 93–112.
Xie, M., Tian, Z., 2018. A review on pipeline integrity management utilizing in-line inspection data. Eng. Fail. Anal. 92, 222–239.
Xu, W.-Z., Li, C.B., Choung, J., Lee, J.-M., 2017. Corroded pipeline failure analysis using artificial neural network scheme. Adv. Eng. Softw. 112, 255–266.
Xu, D., et al., 2023. Failure analysis and control of natural gas pipelines under excavation impact based on machine learning scheme. Int. J. Press. Vessels Pip. 201, 104870.
Yang, S., Fan, J., Zhang, L., Sun, B., 2021a. Performance prediction of erosion in elbows for slurry flow under high internal pressure. Tribol. Int. 157, 106879.
Yang, D., Hou, N., Lu, J., Ji, D., 2022. Novel leakage detection by ensemble 1DCNN-VAPSO-SVM in oil and gas pipeline systems. Appl. Soft Comput. 115, 108212.
Yang, Y., Li, Y., Zhang, T., Zhou, Y., Zhang, H., 2021c. Early safety warnings for long-distance pipelines: A distributed optical fiber sensor machine learning approach. In: Proceedings of the AAAI Conference on Artificial Intelligence. pp. 14991–14999.
Yang, S., Zhang, L., Fan, J., Sun, B., 2021b. Experimental study on erosion behavior of fracturing pipeline involving tensile stress and erosion prediction using random forest regression. J. Natl. Gas Sci. Eng. 87, 103760.
Yang, Y., Zhang, H., Li, Y., 2021d. Pipeline safety early warning by multifeature-fusion CNN and LightGBM analysis of signals from distributed optical fiber sensors. IEEE Trans. Instrum. Meas. 70, 1–13.
Yao, J., Liang, W., Xiong, J., 2022. Novel intelligent diagnosis method of oil and gas pipeline defects with transfer deep learning and feature fusion. Int. J. Press. Vessels Pip. 200, 104781.
Yin, H., et al., 2021. An integrated framework for criticality evaluation of oil & gas pipelines based on fuzzy logic inference and machine learning. J. Natl. Gas Sci. Eng. 96, 104264.
Zakikhani, K., Nasiri, F., Zayed, T., 2020a. A review of failure prediction models for oil and gas pipelines. J. Pipeline Syst. Eng. Pract. 11 (1), 03119001.
Zakikhani, K., Zayed, T., Abdrabou, B., Senouci, A., 2020b. Modeling failure of oil pipelines. J. Perform. Constr. Facil. 34 (1), 04019088.
Zaman, D., Tiwari, M.K., Gupta, A.K., Sen, D., 2020. A review of leakage detection strategies for pressurised pipeline in steady-state. Eng. Fail. Anal. 109, 104264.
Zhang, M., et al., 2022. Defect identification for oil and gas pipeline safety based on autonomous deep learning network. Comput. Commun. 195, 14–26.
Zheng, J., et al., 2022. Deeppipe: A deep-learning method for anomaly detection of multi-product pipelines. Energy 259, 125025.
Zhou, M., et al., 2021. A pipeline leak detection and localization approach based on ensemble TL1DCNN. IEEE Access 9, 47565–47578.
Zhou, D., et al., 2022. Dynamic simulation of natural gas pipeline network based on interpretable machine learning model. Energy 253, 124068.
Zuo, Z., et al., 2022. A semi-supervised leakage detection method driven by multivariate time series for natural gas gathering pipeline. Process Saf. Environ. Prot. 164, 468–478.