

Vol. 14, No. 3, June, 2024, pp. 773-786

Journal homepage: http://iieta.org/journals/ijsse

# Oil and Gas Pipelines Leakage Detection Approaches: A Systematic Review of Literature

Sumayh S. Aljameel<sup>1</sup><sup>1</sup>, Dina A. Alabbad<sup>2</sup>, Dorieh Alomari<sup>2</sup>, Razan Alzannan<sup>2</sup>, Shatha Alismail<sup>2</sup>, Aljwharah Alkhudir<sup>2</sup>, Fatimah Khawaher<sup>2</sup>, Fatimah Aljubran<sup>2</sup>, Atta-ur Rahman<sup>1</sup>



<sup>1</sup>Department of Computer Science, College of Computer Science and Information Technology, Imam Abdulrahman Bin Faisal University, P.O. Box 1982, Dammam 31441, Saudi Arabia

<sup>2</sup> Department of Computer Engineering, College of Computer Science and Information Technology, Imam Abdulrahman Bin Faisal University, P.O. Box 1982, Dammam 31441, Saudi Arabia

# Corresponding Author Email: aaurrahman@iau.edu.sa

Copyright: ©2024 The authors. This article is published by IIETA and is licensed under the CC BY 4.0 license (http://creativecommons.org/licenses/by/4.0/).

https://doi.org/10.18280/ijsse.140310

ABSTRACT

Received: 31 January 2024 Revised: 10 March 2024 Accepted: 22 March 2024 Available online: 24 June 2024

#### Keywords:

artificial intelligence, deep learning, machine learning, pipeline leakage detection, software hardware methods

In terms of its significance, the oil & gas industry is ranked among the top global industries. Like any other industry, it also faces various problems, such as leakage of oil and gas pipelines. The detection of such leakage in the pipelines is essential for an industry or plant to operate properly and maintain environmental safety as well as to minimize its supply-chain losses. The undergoing study systematically reviews the literature comprising of more than a decade (2010-2021) span to summarize the systems, methods and techniques used in oil & gas pipeline leakage detection. Likewise, this paper investigates effective and low-cost leakage detection systems for oil & gas pipelines with their pros and cons. The existing methods are classified into three categories based on their technical characteristics, named as hardware-based (where some hardware is deployed for monitoring the leakage), software-based (where some software is deployed for monitoring the leakage), and intelligent based (where some intelligent predictive algorithm is deployed for the leakage detection) techniques. Each technique was reviewed according to the datasets used, preprocessing techniques (mainly used in the intelligent based techniques that are based on the imagery such as image preprocessing largely like image enhancement, image denoising and filtering), investigated classifiers' efficiencies, results, and limitations. A comparative analysis was conducted to help determine which technology is best for a given operational environment, that is software, hardware, intelligent or hybrid. Further, the paper highlights the gaps in the research and unresolved concerns regarding the development of dependable pipeline leak detection systems and suggests possible research directions to mitigate it.

# 1. INTRODUCTION

In their daily lives, people must strive to preserve and improve the quality of the environment to achieve a healthier life for humanity. However, in recent years, oil & gas pipeline leaks, which are often caused by environmental effects, harmful human activities, or construction near pipelines, threaten the environment. Due to their harmful effects that may disturb the ecosystem as well as human lives and natural resources, and by taking precautionary measures, can be avoided.

Further, the intelligent techniques for predictive maintenance could potentially alarm beforehand where the gross losses in terms of human lives, natural disaster and financial crises can be prevented. Moreover, this is not the only problem, another aspect is the potential financial losses incurred to the oil & gas industry annually. In terms of statistics, alone in United States during 2010-2019, almost 5500 incidents happened. According to the survey there were more than 125 casualties, more than 600 injuries, more than

800 fires, more than 300 explosions, more than \$4 billion financial losses and more than 30,000 people evacuated [1]. In addition, the survey reveals that not all the incidents are reported, and actual statistics are even deteriorating [1].

This paper reviews and summarizes the results and findings of several scientific papers published during 2010-2021, that address the issue of gas or oil pipeline leaks using various techniques. One of the main objectives and motivation is to help the researchers identify the best model to build to address pipeline leakage and provide them with possible methods to solve this problem while considering the pros and cons, beforehand. In this paper, we first provide background on the oil & gas industry in Saudi Arabia, the leak issues that occur in them, and the techniques that help solve this problem, such as machine learning techniques etc. Next, we review the literature on techniques that have been used for early pipeline leak detection. Then, this paper focuses on the models, methods, techniques, and approaches used in each study regarding the datasets nature, preprocessing methods, extracted features, investigated classifiers, and the results. Finally, we present a table containing the details of each study we reviewed; the summaries are presented and placed in four categories: hardware-based, software-based, intelligent-based, and other miscellaneous smart approaches to pipeline leak detection. Each category has its own pros and cons, and its effectiveness varies depending on the operational environment and provided nature of dataset.

Respectively, in the hardware-based approaches, some hardware components like sensors etc. are deployed for leakage detection. Similarly, in software-based approaches, software tools/packages are used as monitoring systems like calibration tools etc. In the intelligent-based approaches, predictive maintenance techniques are mainly deployed, that intuitively forecast the potential leakage. In the hybrid approaches, more than one type of technology is used in addition to the latest drones and surveillance technologies [2].

From an oil & gas company's perspective one technique may be better than the other or sometime a hybrid approach can be more useful. For instance, collaboration of hardware and software-based techniques may result in better hardware handling and reporting by means of the applied software. Similarly, software-based schemes can be further fine-tuned with the help of intelligent machine learning approaches. Likewise, one or more than one approaches can utilize smart techniques like surveillance using drones etc. to enhance their effectiveness more. Other factors for consideration are the industry's region/location, environmental factors, field radius and span, available resources, hardware (pipeline) life and material, its suitability to the ambiance temperature etc. and the sanctioned budget for the employment of the surveillance techniques for leakage detection [3].

The major motivation behind the study is that Saudi Arabia is an oil rich Kingdom and though there are other factors, but its gross domestic product (GDP) is mainly depending upon the oil and gas sector. There are several state-owned as well as private organizations working in this regard rigorously. The undergoing study is a potential contribution towards the supportive side of this industry in terms of predictive maintenance approaches potentially helpful for leakage detection to minimize the potential losses in terms of capital as well as human lives. In this regard, the current study surveyed several approaches used in literature to comprehend the issue and find potential areas of research and improvement. Further, the major achievements and limitations of the existing studies have been enlisted and research challenges, directions and opportunities are stated based on the review.

The sectioning of the paper is as: section 2 briefs about the oil and gas industry and essence of AI and ML in this field. Section 3 provides the systematic literature review of more than a decade period organized by category; section 4 discusses the reviewed literature and concludes the paper.

#### 2. BACKGROUND

The oil & gas industry is one of the most important industries generally around the globe including Europe, Asia, and America etc. and septically in Saudi Arabia. Several countries of the world are dependent heavily on such natural resources. In this regard, the related industries always remain in the search of new resources and deploy several teams in the expected areas. Oil and gas reservoirs are geographically situated in each other's vicinity. Countries being rich in such natural resources are privileged compared to those lacking them.

It is a well-known fact that Saudi Arabia's main GDP relies on the oil & gas industry. Several renowned organizations and industries have been working day and night in the fields. For instance, Saudi Aramco, Weather Ford, Schlumberger and many other. Experts, engineers, and workers are hired from all around the globe and around 10 million barrels of crude oil has been extracted on daily basis [3, 4]. Recently, several technologies have been used to solve common problems in the oil & gas industry. One of the most critical problems is oil & gas pipeline leaks because it has a significant impact on the environment and wastes resources. Nonetheless, currently, the fear of risks is beginning to fade with the advancement of technology [4]. The early detection of pipeline leaks will solve several problems and may increase safety measures taken to protect a company's employees. The main objectives of the pipeline leakage detection system are to save lives, reduce monetary losses, detect pipeline leaks before complications occur, and develop a highly accurate artificial intelligence (AI) based systems, more specifically using machine learning (ML) and deep learning (DL) based models [5, 6]. ML is a branch of AI, which is a statistical technique that learns from a set of inputs and outputs (examples) to perform tasks that require human intelligence. Based on the problem type, the ML model could be supervised learning, unsupervised learning, semisupervised learning, or reinforcement learning. It further encompasses transfer learning and ensemble learning approaches [7, 8]. ML is used in diverse fields, and it is widely used in prediction and early detection applications. ML models are trained in three main processes: decision, error function, and model optimization. In the decision process, the model feeds the data and attempts to learn its pattern to make a classification or prediction. The error function evaluates the model's performance. Finally, the optimization process updates the weights to reduce the difference between the target and predicted output [7]. AI has shown a significant impact in many fields, and one of those fields is the oil & gas field. As the oil & gas industry is one of the most important resources in the world, and many companies are working in this field, AI can be used and implemented to enhance the industry in this field. Like other industries, the oil & gas industry is facing many problems and difficulties that need to be solved and AI can provide. the needed solutions [8, 9]. AI can enhance the exploration process and make it easier for companies to find which areas to work in. Also, AI plays an important role in the production process and enables the company to work on a specific timeline, with the ability to predict coming problems and prepare for them in advance [10]. Also, AI can solve one of the greatest problems in the oil & gas industry, which is pipeline leakage. AI can help in the early prediction and detection of pipeline leakage and to specify the exact location and size of leakage. Although smart detection techniques are useful in their pursuit yet offer a significant investment and regular follow ups [2].

# **3. SYSTEMATIC LITERATURE REVIEW**

This research presents the literature review of more than forty state-of-the-art studies in pipeline leakage detection. The review is divided into three subsections based on nature of the techniques, namely, hard-ware-based, software-based, and intelligent-based techniques. The stats of the reviewed studies are given in Figure 1 and Figure 2, respectively.

Moreover, Figure 3 presents the taxonomy of the reviewed methods, approaches, and techniques.

#### 3.1 Hardware-based techniques

This section explores the hardware-based techniques that use devices to monitor the external parts of pipelines to detect leakage. These leak detection methods require physical contact between the pipeline being monitored and the sensors. The methods reviewed in the literature include the resistivity method, pyroelectric infrared sensors, wireless sensor networks (WSN), and the acoustic signature method.

In their study, Somov et al. [11] used WSNs to detect methane leakage in boiler facilities. Wireless sensor networks have a network coordinator that controls nine battery-powered wireless sensors. Received signal strength indicator and link quality indicator metrics were used to assess the wireless links. The researchers' goals were to obtain faster and more accurate responses from the sensors and to deliver methane leakage data from the sensor to the network coordinator. There were difficulties in the experiment securing reliable communication channels between the network coordinator and the sensor nodes. To remedy this issue, the researchers used 2dB sensor nodes and 5dB external antennas. As a result of the wireless link evaluation experiments, both the received signal strength indicator and the link quality indicator metrics were considered to achieve high-quality wireless links. When the received signal strength indicator was greater than -79.3 dBm, the packet delivery rate was higher than 80%, and it reached 100% when the link quality indicator was around 210. Temperature measurement was a key factor in achieving accurate analysis of the sensors' data. About 10,000 data packets were sent within 20 seconds, which proves the sensors' speed. For future research, the researchers plan to study the level of methane concentration, sensor response time, and battery status, which were collected in the experiment.

Similarly, Sun et al. [12] designed a system that detects gas leakage in a buried pipeline. The system was designed based on the resistivity method. In this experiment, geoelectric models of gas leakage were constructed in various states. The system used large-scale theoretical calculation of resistivity detection images. Then, the results were compared with the results of a small-scale outdoor gas leak simulation experiment. However, the geological conditions of the buried pipelines in the actual application are challenging and still require improvement. For future study, the natural electric field could monitor the micro-fissure of soil, and the law of self-potential fluctuation could be used to predict the failure state of the soil caused by gas shock. The results from this exterminate verified that the resistivity method is feasible in detecting gas leakage in buried pipelines. Furthermore, in this paper, Erden et al. [13] proposed a novel method for detecting and monitoring volatile organic compound gas using pyroelectric (or passive) infrared sensors that are widely used in practice to detect motion. These sensors were used because they do not require vapor to reach the sensor; it is sufficient for the gas to be in the viewing range of the pyro-electric infrared device. Additionally, they used Markov models to identify and analyze analog signals in real-time. The wavelet coefficients were used by Markov models as feature parameters. Consequently, based on the experiments, they concluded that using a pyroelectric infrared sensor without a Fresnel lens on it enabled them to sense gas leaks at up to 3 m. Conversely, they proved that Markov model use is optimal to detect volatile organic compound gas and process the wavelet transformed sensor data.







Figure 2. Studies classification in 2010-2021



Figure 3. Taxonomy of the oil & gas pipeline leakage detection approaches

In addition, Kusriyanto et al. [14] developed a device monitoring system and early detection of liquid propane gas (LPG) gas leakage using WSN. To monitor and detect gas leaks early, they used gas sensors of the MQ-4 type and the AVR microcontroller family to control the detecting device. Furthermore, XBee PRO S2B workable devices were integrated into the system as interfaces for the wireless networking system and are used to transmit sensor data from the detection points to an integrated computer and visual basic software program at a monitoring center. Moreover, the authors considered it important to note the occurrence of LPG leaks, reported by an alarm sent through GTalk. Moreover, the system provides an early warning bell. As a result, they designed an early detection system for LPG leakage with WSNs, using the line equation y = 1.000x + 0.004 with a 3.46% error, where the MQ-4 gas sensor output is linear and proportional to the change in the input.

Similarly, Khan [15] proposed an automatic gas leakage detection, alert, and control system. The proposed system was able to detect any gas level changes in the air, identifying if it was above the safety level and sending an alert to the user. The main components of the proposed system were Arduino UNO R3, and a semiconductor MQ-6 gas sensor. The MQ-6 gas sensor could detect several gas types, such as LPG, butane, propane, and natural gas. The proposed system demonstrated adequate performance in detecting and controlling gas leakage within two seconds, and it is also cost-efficient and can be used easily in homes. The proposed system could be enhanced to calculate the amount of wasted gas and gas usage. Moreover, intelligent techniques could be applied to the proposed system to improve its capabilities.

Similarly, Wang et al. [16] aimed to design an in-pipe detector that is suitable for urban gas pipeline leak detection. They used several methodologies: power spectral density, artificial neural networks (ANN), and Fourier transformation. The data set was collected during the traveling process in the pipeline using the detector. The validation results showed that the leakage can be detected with a precise of 96.87% by the trained ANN model. The proposed model detected gas pipeline leakage with high accuracy. In the future, the researchers could acquire more data to improve the precision of the leak recognition. In addition, Yan and Rahayu [17]

designed and developed a gas leakage monitoring system. The combustible gas sensor (MQ9) was used to detect carbon monoxide gas (CO) and methane gas (CH4). However, the detection range was between 200 to 10000ppm. The Arduino Uno was used as a microcontroller for their system. Whereas Zigbee was used to send the gas sensor readings to a specific monitoring system that displays a Graphical User Interface (GUI) on LabVIEW. Users benefit from the information and take immediate action; otherwise, the system and gas supply shut down automatically within 10 minutes to prevent serious conditions.

In their research, Guo et al. [18] proposed to develop a system based on mobile WSNs to monitor gas leaks and send early warnings in the event of any leakage. The system consisted of two parts: remote sensors and an analytical server. The sensor is the sensing station, equipped with sensor terminals that are either fixed or mobile. The sensor terminal consists of several elements integrated and used during the field data collection process. These elements are the gas sensor, control unit, power unit, GPS receiver, and GPRS unit. The analytical server receives the data from the GPRS unit via NetAssist, and after processing, it stores it inside the MySQL database. Additionally, a cloud platform was created to display the data of each sensor, such as the sensor number and its location. Moreover, this cloud platform can be used to monitor real-time sensor readings of gas concentrations. If the concentration exceeds the warning threshold, the analytical server analyzes the suspected leak site and sends a message, alerting the concerned authorities. The sensor readings of gas concentrations in the area were classified into three states: sleep mode, wake mode, and transmit mode. The sleep mode is activated with a periodic check every 30 min if the concentration is less than 3 ppm, and the wake mode is activated with a periodic check every 10 min if the concentration is between 3-10ppm. Transmit mode is activated with a periodic check every min if the concentration exceeds 10 ppm; if the transmit mode is activated, it indicates that there is a gas leak in the area. After conducting several field experiments, the results indicated that the system developed in this study is reliable and practical. Moreover, it can monitor leakage in real-time and send early warnings in the event of an emergency.

Study	Year	Methods and Equipment	Results
Erden et al.	2010	Pyro-electric Infrared (PIR) sensor, and	Using a PIR sensor without a Fresnel lens on it, they were able to
[13]		Markov models (MM)	record VOC gas leaks at up to 3 meters.
Somov et al.	2012	WSN, Received signal strength indicator	When the RSSI is greater than -79.3 dBm, the packet delivery rate
[11]		(RSSI) and link quality indicator (LQI)	(PDR) is higher than 80% and it reaches 100% when the LQI is
		metrics.	around 210.
Yan and	2014	Gas sensor (MQ9) was used to detect carbon	A detection range between 200 to 10000ppm.
Rahayu [17]		monoxide gas (CO) and methane gas (CH4).	
Kusriyanto et	2018	WSN, gas sensor MQ-4	They concluded using the line equation $y = 1.000x + 0.004$ with a
al. [14]			3.46% error, where the MQ-4 output is linear and proportional to
			the change in the input.
Bolotina et al.	2018	Acoustic phased array antennas	Ability to find and locate a leakage of 25 l/hour from 50m.
[19]			
Guo et al. [18]	2019	Mobile WSN	Gas leakage results when concentration exceeds 10 ppm in relation
			to real time.
Khan [15]	2020	Arduino UNO R3 and MQ-6 gas sensor	Detection and preventing gas detection within 2 seconds.
Sun et al. [12]	2021	Improved negative pressure wave method	Leakage position has an absolute error of 0.38 m.
		based on FBG based strain sensors and	
		wavelet analysis.	
Wang et al.	2021	ANN, Fast Fourier transform (FFT) and	Accuracy of 96.87%
[16]		PSD.	

Table 1. Summary of hardware-based techniques

Likewise, Spachos et al. [20] proposed a paper aimed to develop a system for monitoring and locating gas leaks in indoor environments using a WSN. The proposed system consists of three modules: fixed nodes, mobile nodes, and a control room. Three fixed nodes are distributed in the form of a triangle in a fixed location for monitoring and placing a mobile node moving through those fixed nodes to collect the data monitored by each of the fixed distributed nodes and then send it to the control room. The control room studies and analyzes the collected data to determine if there is a leak and locates that leak by the mobile node's location. The results showed the effectiveness and flexibility of the model, as it can be deployed in several environments, buildings, hospitals, mining tunnels, and commercial centers. A summary of hardware techniques is given in Table 1.

### 3.2 Software-based techniques

This section discusses the software-based techniques that rely on internal fluid measurements to monitor parameters. such as density, temperature, and pressure, related to the oil & gas flow inside pipelines that can indicate leakage. The methods in the literature include negative pressure wave and IoT. Hou et al. [21] designed a system to detect gas pipeline leakage using an improved negative pressure wave method based on Fiber Bragg grating based strain sensors and wavelet analysis. To enhance the method, the researchers incorporated the natural gas velocities and variation of the negative pressure wave into the negative pressure wave leak location formula, compound Simpson formula, and dichotomy identification to solve this modified formula. Furthermore, they used an FBG based strain sensor to collect the negative pressure wave signals to overcome the difficulty of installing traditional pressure sensors. This method provides many beneficial features, such as ease of insulation and low cost. Moreover, a wavelet transform-based method was used to locate the pressure drop points within the FBG signals. The results from this study indicate that the method accurately located the position of a natural gas pipeline leak. The calculated leakage position has an absolute error of 0.38m. Further, Jiang et al. [22] introduced a gas pipeline anti damage and early warning monitoring system. Sound and vibration detection techniques combined with IoT sensing technology were used to construct this system. The system monitors abnormal states of the gas pipeline and transmits an alarm through the data analysis platform. The noise from the pipelines is monitored by sound detection technology, and through the IoT, the detected vibration is sent to the data analysis platform. The frequency and amplitude of the pipeline leakage are sent to the maintenance personnel in the data analysis platform, which determines whether the gas pipeline is leaking. This system is designed to achieve real-time, early warning pipeline monitoring.

Additionally, Sharma et al. [23] designed a gas leakage system by applying embedded systems and the IoT. The system aims to reduce accidents and promote safety using the existing electronics and technology. The gas detection and alarm system will not only alert about the leakage but also hold a unique selling point (USP) of automatic gas shutting feature to control any leakage. Furthermore, the system alerts the users by alarm and message using the IoT. The addition of basic small inputs, such as temperature, is used to monitor the temperature of the cylinder. IoT is amplified by Android and GSM for sending emails. Additionally, light-emitting diodes (LED) are used to increase the warning system's reliability through flashing gas leak indicators and sensors. The gas system proved to be a valuable safety tool since it was able to perform all four of the functions mentioned.

Ralevski and Stojkoska [24] aimed to detect gas leakage in houses based on the IoT system, which consists of one laptop and two Raspberry Pi's equipped with sensors. The proposed system could work in hazardous environments. The devices collect data to enable environmental conditions analysis. Gas leakage is detected by measuring gas concentrations and temperature. The moving average (MA) algorithm merged with the IoT distributed system was used to assure low power consumption. The system depends on some defined rules that enable it to communicate with specified devices for detection. Time series forecasting approach is used to minimize the measuring node sent packages and enhance the communication process. In case of leakage detection, the system will inform the end-user by sending a notification to their mobile device. The researchers evaluated the MA algorithm that was used in this system by making comparisons between several types of algorithms. MA provides accurate prediction of indoor temperature measurements. In future research, statistical analysis of the measured temperature textual file could be used to enhance the prediction algorithm.

Moreover, Debnath et al. [25] proposed a low-cost gas leakage detection and warning system using IoT. In case of gas leakage or temperature increase, the proposed system was able to call users, alert them, and send the graphical location to the webserver. The proposed system was built using multiple hardware and software components, and an ML model was used to measure the system performance accuracy. The hardware components of the proposed system were an MQ-2 sensor to detect gas and smoke, an IR sensor to detect light, a GSM module for mobile communication over the network, a mini-node microcontroller unit (Node MCU), a buzzer for sound alerts, and a DHT11 temperature sensor. By using these components, the proposed system was able to sense gas leakage, fire, and high temperature. To implement the software, an Arduino microcontroller code was used. To check the proposed model's accuracy and performance, several ML algorithms were used, such as multilayer perceptron (MLP), support vector machine (SVM), and Naïve Bayes (NB). The proposed model exhibited an average prediction accuracy of 92.6275%. However, the proposed system can call one user in case of an emergency. More work should be done to support multiuser alerts handling.

Similarly, the study of Alshammari and Chughtai [26] aims to introduce and design a gas leakage monitoring system with the help of the Internet of Things (IoT). The information captured by the gas sensor (MQ-5) was posted into the cloud. The detection of gas leakage was done under most atmospheric conditions. Also, Arduino (UNO-1) was used as a central processing unit to control all the components. However, when the sensor detects gas leakage, an alarm is raised in the form of a buzzer. Moreover, to display the gas leakage location, alert the users, and turn on the exhaust fan to suck out the leaked gas in a specific location, the alarm was supported by LCD display. After the implementation of the design, the device was able to detect the gas leakage accurately and generate a message to be transmitted to alert the users. PCB gas leakage detector was used to obtain the practical results, up to10000ppm concentration was detected by a high accuracy sensor and LED turns red. However, with a small modification of the system, the system could be efficiently used for household purposes to detect gas leakages.

 Table 2. Summary of software based techniques

~ -			
Study	Year	Methods and Equipment	Results
Hou et al. [20]	2013	Improved negative pressure wave method based on FBG based strain sensors and wavelet analysis.	Leakage position has an absolute error of 0.38m
Sharma et al. [22]	2018	IoT, USP	The system can perform all the functions and a valuable safety tool
Ralevski et al. [23]	2018	MA algorithm merged with the IoT system and time series forecasting approach.	MA algorithm poofs its efficiency by producing a good prediction in indoor temperature measurements.
Debnath et al. [24]	2020	MQ-2 and IR sensor, GSM module, Buzzer, DHT11, Arduino, ML	Accuracy of 92.6275% for sensors predictions.
Alshammari and Chughtai [25]	2020	IoT, Arduino (UNO-1), gas sensor (MQ-5)	Up to10000ppm concentration detected by a high accuracy sensor.
Swetha and Shwetha [26]	2021	MQ-2 gas sensor, IoT	Detection of LPG/CNG gas from 200 to 10,000ppm.
Jiang et al. [21]	2021	Sound and vibration detection techniques combined with IoT	System designed for real-time early warning and monitoring of pipeline.

Swetha and Swetha [27] developed a system for controlling and detecting gas leakages using MQ-2 gas sensor. The system was programmed with embedded C to evaluate the sensed information. When it is greater than a specific threshold, an alert SMS will be sent to the users to activate the servo motor and turn off the gas valve. However, the system was able to detect LPG/CNG gas concentrations range 200-10000ppm. The device is portable, low-cost, efficient, user-friendly, lightweight, safe, and easy to detect gas leaks. Table 2 shows a summary of software-based schemes.

#### 3.3 Intelligent-based techniques

This section reviews the use of AI and data processing techniques to detect various leaks. With the help of AI systems, people can overcome many risks by making faster decisions. However, with ML and deep learning, AI-based systems can be developed to help people solve overly complex and repetitive problems because AI systems ensure 24/7 service to reduce workers' stress and improve work efficiency.

Babu et al. [28] designed a smart natural gas leakage detection system for households. Several incidents have been reported over the year due to gas leakage in India. The suggested gas leakage detection system chains innovative sensor equipment, real-time supervising, and automated alert systems to safeguard, well-timed detection and reaction to gas leakages. The MQ2 sensor assists in recognition of gas leakage and exhibits the competence to identify a comprehensive selection of gases, incorporating methane, propane, carbon monoxide, and hydrogen. The accumulated data is investigated using advanced algorithms to differentiate between normal and leakage surroundings gas readings. The proposed NodeMCU equipment, together with its Wi-Fi abilities, acts as the main system's controller. It collects data in the real time and sends it to the edge server for processing.

Hubert and Padovese [29] used other ML algorithms to develop a model for early gas leakage detection in underwater pipes using passive acoustic emission (AE), which can indicate leakage. The data set used for this research was from a pilot study with simulated leakages that had a total of 1,900 seconds of recordings. The algorithms used to build the classifier- detectors are the random forest (RF)s algorithm and the gradient boosting (GB) trees algorithm. The results of the study suggested that GB tree algorithm was the best fit, achieving an accuracy of 81% using 5-fold cross-validation. For future research, the authors intend to investigate other classification strategies. In addition, they are conducting new experiments to enrich their data set. Based on their results, the only limitation was the accuracy obtained because it is possible to obtain a higher one. Therefore, to improve their accuracy, they intend to examine other classification methods. Chi et al. [30] aimed to identify the best ML method that addresses pipeline leaks. They worked with an experimental data set of 130 experiments collected from several states. The ML methods used in this paper were RF, SVM, ANN, KNN, and decision tree (DT). The results indicated that the RF classifier was the best model for leakage detection; it outperformed other ML methods with a classification accuracy of 88.33%. For future method development, the authors will conduct more experiments to test the practicality of this method and confirm its efficiency. The limitation in their research was related to the data set size and the achieved accuracy. Since the data set is considered small, increasing its size may improve the accuracy rate.

Kampelopoulos et al. [31] used various ML algorithms for monitoring system to detect potential leaks based on variances from a pipe's normal operation noise. They used a data set that had pipe noise measurements from typical operating conditions and artificial leak measurements. In this data set, data was collected at intervals of 1 sec, and measurements were taken for approximately four hours. The ML models used to build the proposed system were SVM and DT. These two classifiers were trained, tested, and compared with each other. As a result, DT was found to be a better fit with 97.9% accuracy than SVM with 97.1% accuracy. In this study, the model must generate the least possible number of false alarms (false leak predictions). Thus, the model with the highest recall score is the most appropriate. In this case, the SVM classifier achieved the highest noise recall (97.07%) while the DT achieved 95.83%. The study was limited by false alarms, and to overcome this, the researchers should train the algorithms for accurate detection and provide more signal features to the existing ones.

In their study, da Cruz et al. [32] aimed to detect and locate gas pipeline leaks using ML techniques and acoustic sensors. Additionally, the researchers' aims included solving two major problems: detecting small leakages on pipelines that operate at low pressures and reducing false alarms caused by external noise. To perform this experiment, an experimental apparatus was built in the Process Control and Automation Laboratory at the Chemical Engineering Facility at UNICAMP to collect the required data set for training ML models. The data set was collected using four microphones, and the total number of instances was 1,800,000 instances. To convert the data into the suitable frequency domain, the researchers used the fast Fourier transform. The algorithms included in this study were logistic regression, KNN, SVM with a linear kernel, SVM with a radial basis kernel, random for adaptive boosting, and extreme GB (XGBoost). The data set was split into 80% and 20% for training and testing. To evaluate the proposed model, the researchers used detection accuracy and the average error for localization. The RF algorithm resulted in the highest leakage detection accuracy (99.6%), and XGBoost resulted in the lowest average error for localization (1.75%). The localization results indicated lower performance when new samples were used in the model. This means the proposed system may result in poor performance in a real-world scenario.

Narkhede et al. [33] proposed a novel method to detect and find gaseous emissions using multimodal AI fusion techniques. This research was conducted using a data set containing images of gas samples collected manually using a sensor array and a thermal camera. For preprocessing, they used data augmentation techniques to increase the data set size. They also increased the diversity of finite thermal images using data augmentation techniques, such as rescaling and resizing. They applied a convolutional neural network (CNN) to extract features from thermal images while using the long short-term memory framework to extract features from sequences of gas sensor measurements. However, they applied the fusion model after noticing that applying each model separately produced weak results. Thus, they obtained an accuracy of 96% for the fusion model compared to 82% for an individual model using the long short-term memory framework, and 93% for a CNN model. The fusion of multiple sensors and modalities outperforms the output of a single sensor. Due to the large amount of training data needed for effective operation, this study faced a data set limitation. Thus, they took advantage of data augmentation techniques to increase the data set size and overcome the limitation. Oliveira et al. [34] aimed to detect pipeline leaks using a set of ML techniques. The data was collected from sensors' signals, and then the plant historian was used to convert the data into understandable data. Data mining techniques were used to interpret, clean, and preprocess the data. The pipeline energy balance was monitored using an anomaly detection approach and a linear regression ML model to detect pipeline leakage. The system treats any outliers that are detected as a leak. The largest challenge was to reduce false alarms by adjusting the threshold that classifies behavior as normal or abnormal. As a result, the proposed system detects pipeline leaks with 3-5 wrong alarms per month. In future studies, the researchers want to increase the leakage data and use the wavelet package algorithm to accurately locate leaks. To prove the high performance of this system, the researchers need to mention the accuracy of the proposed system.

Akinsete and Oshingbesan [35] proposed a system to detect gas pipeline leakage using intelligent models and data analytics. A leak detection algorithm acts as a classifier, which was applied with five intelligent models to act as regressors: RF, GB, SVM, DT, and ANN. The models were turned by grid search using the mean average error, root-mean-square error, and the coefficient of determination. The data set was obtained from the Supervisory Control and Data Acquisition measurements of operational data with 80% and 20% split for training and testing, respectively. The trade-off between accuracy and reliability was a challenge that led to the reduction of performance accuracy to increase reliability. As a result, all models are reliable. The SVM and ANN models have the highest accuracy, which is 98%. The RF and DT models have the highest sensitivity because of their ability to detect 0.1% leaks, and the models' performance is suitable compared with the real-time transient model. In future experiments, data analytics, big data, and artificial intelligence tools could be used to enhance the detection results. The models could be used with the real-time transient model to enhance the detection process and implement the models with little or no operational data.

Wang et al. [36] proposed an ML model to detect methane emissions from oil sites. The data set used comprises field measurement data on emissions and general on-site demographic data, generated with a gas-optical camera for leak detection and combined with a tool for determining emissions rates. The data was collected from 436 sites, including 229 oil production sites and 207 gas production sites. A training set and a test set were constructed using a 75% to 25% split. The proposed system predicted the locations with the highest emission so that operators could be guided through the system. In the model approach, the researchers used the marginal return of emission coverage to identify the size of cutoff emission. The algorithms used to build the model are logistic regression, DT, and adaptive boosting. These algorithms were selected because they do not require engineering many features for the data set. During the model performance evaluation, the authors ran three scenarios. The first scenario involved surveying each of the production sites in the data set in random order. This scenario was implemented in LDAR methane regulations in Canada and the United States. second scenario involved extracting emission The probabilities from the model, classifying them from top to bottom, and guiding operators. The third scenario included comparing the amount of gas production between sites and arranging the data from the highest to lowest production to direct operators to the areas with the highest increase in production; these sites have an increased possibility of leakage. The results demonstrated that logistic regression had the best performance with 70% accuracy, the highest recall and sensitivity rate (57%), and a balanced accuracy rate (66%). During the evaluation of the proposed model, the balanced accuracy rate (the average for each category), the accuracy rate, and the recall and sensitivity rate were used. In this study, the researchers faced difficulties in detecting methane emissions and predicting the release of methane due to its random and variable nature. For future study, an improved version of this model could include a significant amount of emission data from a larger geographic scale, incorporate more attributes, and consider the return on investment from methane mitigation.

Rashid et al. [37] proposed a distributed system to monitor and detect leakage in oil or gas pipelines using a WSN supported by ML techniques. The data set comprised raw data from individual sensor nodes. A training set and test set were constructed using a 60% to 40% split. In the pre-processing stage, the noise was reduced from the data using a low-pass filter and Daubechies wavelet transform. The algorithms used to build the model were GMM, KNN, and SVM. During the feature identification and reduction phase, nine features were selected to improve classification performance, and the selection process took place after two tests: Wilcoxon and Ansari-Bradley Network (RNN). The SVM model was the best performing model with the highest leak detection accuracy (94.73%) and leakage size estimation (92.3%). In the study, the researchers faced limitations in identifying slow and small leaks within the pipeline, so the situation required an algorithm capable of accurately identifying minor leaks.

While other researchers solely used the SVM algorithm, in their research, Ahn et al. [38] proposed to use ML models using AE signals, which reduce signal noise, to improve early leakage detection. The proposed model included using principal component analysis for signal pre-processing and a genetic algorithm for feature selection. For the proposed model, the researcher collected signals from AE sensors. After applying the techniques for feature selection, 30 features had been used in training the proposed model. The SVM algorithm used in this study resulted in 100% performance of vibration signals, while AE signal was 80% and genetic algorithm and principal component analysis were 70%. However, using other ML algorithms could result in better performance of AE signals, genetic algorithm, and principal component analysis.

Xiao et al. [39] proposed a system to detect gas pipeline leakage through acoustic signals and an SVM. The relief-F algorithm is used to select the best features and input them into the SVM algorithm to determine the intensity of a leak. The researchers used a data acquisition system where 75% of the data is used for training and 25% is used for testing. In the preprocessing stage, the researchers selected the best wavelet through a wavelet entropy-based algorithm. The signals' noise was removed using the universal threshold rule and extracting useful leak features. As a result, when the system uses the three best features, its accuracy reached 99.4%, which can determine a leak or non-leak state. When the system uses the five best features, its accuracy is reduced to 95.6%, which can classify the intensity of leaks and normal states.

Liu et al. [40] designed a leakage detection method for water pipelines. The algorithms used to build the proposed model are based on ML, WSN, and SVM. The model was implemented on 100 data sets of non-leakage and leakage signals. The wireless sensors installed on the pipelines collected the data and utilized a 4G network to remotely transform the data. The complexity of the network architecture, the large scale of the water pipeline networks, and the environmental conditions could be challenging for this proposed system. The results from the proposed model demonstrate the effectiveness of identifying a leak in a water pipeline with lower energy consumption compared to the networking methods used in a conventional WSN. The proposed algorithm achieved 98% classification accuracy. However, this study is limited to detecting water pipeline leakage.

Similarly, Chen et al. [41] used SVM algorithms to build a distributed fibre-optic alarm system. The system maintains the integrity of oil & gas pipelines by isolating a leak site and implementing early warning monitoring-created with the support of one of the machine learning algorithms. The data set comprised the signal data generated by vibrations along the pipelines, which were collected using a distributed optical fiber vibration sensor. One hundred thirty sets of data were collected. One hundred sets were randomly assigned from a large amount of data for each SVM learning field trial procedure. Additionally, 30 data sets for each action were randomly assigned to the SVM test that completed the learning. The SVM algorithm had a 90% accuracy rate in recognition. However, the researchers faced limitations in identifying abnormal events along the pipeline due to multiple classification issues, so it was solved using the "one-to-one" method. Through future studies and field experiments, the model will be upgraded to improve identifying abnormal

events.

Wang et al. [42] aimed to discover an effective method for detecting long-distance oil pipeline leakage and avoid issues that may result from it. The dropout problem is a classification of problems, so based on an ML algorithm, a new classification approach using SAE, lead-follower particle swarm optimization (LFPSO), and SVM has been proposed. The SVE algorithm obtains features from pipeline leak data. The LFPSO algorithm improves the parameters of the SVM algorithm so that the probability of trapping in the local optimum is effectively reduced. The data set comprises 470 sets of pipeline leak data; it consists of 275 sets of normal data and 195 sets of leaked data. These data were randomly divided into 300 training data sets and 170 test data sets. For the preprocessing stage, noise is reduced in the data based on the decomposition of variational mode. The algorithms used to create the model are SVM, SAE-softmax, and BP algorithms. The researchers targeted three prediction methods which are SAE-softmax, SAE-LFPSO-SVM, BP, and SVM. During the experimental process, each experiment was conducted 10 times separately to ensure the effectiveness of the suggested algorithms. The proposed SAE-LFPSO-SVM algorithm outperformed the other algorithms by 92.49% in sensitivity, 100% in positive predictive value, and 96.41% in total classification accuracy. The authors' limitation in this study is the difficulty optimizing the SVM parameters, which was addressed by the LFPSO algorithm.

Banjara et al. [43] proposed the use of AE technology to detect leakage in pipelines through systematic analysis of signal parameters based on some machine learning algorithms. The data set contained AE signals collected using sensors connected to the pipeline. The researchers targeted two prediction methods: binary classification and multiclass classification. The algorithms used to build the model are SVM, and relevance vector machine. In binary classification, the data was divided into two parts: a 171-sample training set and a 20-sample test set. As for the multiclass classification, its data was divided into two parts: a 699-sample training set and a 70-sample test set. After conducting experiments, the results revealed that the model's performance in the binary classification method was better than that of the multiclass classification method. Regarding SVM, it was 99.92% with a misclassification error of 0.0828%, and the relevance vector machine was 98.5% with a misclassification error of 1.5%. One of the limitations that the authors encountered in their research was that the relevance vector machine algorithm only provided satisfactory results when used in binary classification.

Qu et al. [44] applied ML to develop a pipeline leakage detection and pre-warning system. The model could localize the detected leakage. The data set used in this study was collected from sensors during private experiments. SVM was used to perform this experiment. Moreover, accuracy and precision were used to evaluate the proposed model. The model was 95% accurate in leakage detection and had  $\pm 200$  m precision for localization. More powerful ML algorithms can be used to improve performance.

Some researchers, such as Mohamed et al. [45] focused on the use of several types of ANN. They aimed to use ML algorithms to estimate the defect depth in pipelines to determine defect severity. The proposed model includes detecting the length of the defect using pattern-adapted wavelets and applying an ML algorithm to detect the depth of the defect. The researchers used data from magnetic flux leakage sensors to train the proposed model. The collected data contained signals from the magnetic flux leakage sensor readings, and it was typically used in training the model. To process the raw data for training the model, the researchers applied some feature extraction techniques and used five statistical features, which were maximum magnitude, peak-topeak distance, integral of the normalized signal, mean, and standard deviation. The researchers used diverse types of feedforward neural networks (FFNN), which were static FFNN, cascaded FFNN, and dynamic FFNN. To evaluate the performance of the proposed model, the researchers compared the accuracy of each model for various error tolerance levels. The proposed model resulted in more than 86% accuracy for different error tolerance levels using dynamic FFNN. The statistical features used in this study may not be enough to acquire satisfactory results; applying other feature selection techniques and using different statistical features could result in achieving higher accuracy and lower error tolerance levels.

Layouni et al. [46] used ML algorithms to automate the analysis of magnetic flux signals to estimate the length and depth of metal-loss defects. Since the depth of defect has a key role in estimating the severity of a defect, and the magnitude of magnetic flux signals is much higher for larger defects. The researcher collected 1,300 raw magnetic flux signals that include five features to train the proposed model: peek-to-peak distance, mean, standard deviation, and integral of the normalized signal. The algorithm that had been used in this study was the FFNN algorithm. To evaluate the proposed model, the researchers used prediction errors. The proposed model resulted in a satisfactory performance prediction error in the range of  $\pm 9\%$  of the real defect depth. The model was trained using experimental data. Using real data is needed to measure the real performance of the model. The researchers will use other defective shapes for training the model in future studies. Kim et al. [47] aimed to detect the size and location of leakage in subsea natural gas pipelines. The researchers selected the most sensitive variables based on a dynamic model related to detecting leakage and various flow simulations. Furthermore, an ML data set was generated by changing these sensitive variables. The data set was collected in two stages: A high production stage and low production stage. Each stage has a total of 3,200 data sets, and 90% of the data is used for training and validation. The deep neural network method and ML techniques were used to build the proposed model. The mean error and R-square techniques were used to evaluate the model accuracy. The proposed system achieved 80% accuracy, which is considered highly accurate. Additionally, the researchers proposed a flowchart for leak detection in gas pipelines. However, the researchers could use advanced preprocessing techniques to improve the accuracy of their model. While other researchers only used CNN algorithms.

As Song and Li [48] applied different architecture models of CNN and screw connections to simulate gas leakage experiments for leak detection in galvanized steel pipelines. The data set was collected from a sensor placed on the pipeline, and 70% of data was used for training and 30% was used for testing and cross-validation. In the preprocessing stage, the researchers removed noise from the collected signals and put the cleaned signals in fixed-length samples. As a result, the three-layer CNN model with a big kernel reached more than 93% accuracy. The challenge that the researchers faced was that there is little research on galvanized steel pipe systems. Furthermore, the noise of internal gas flow is like gas leakage, which leads to false leak alarms. The researchers should increase the sampling length or rate to enhance accuracy.

Ghorbani and Behzadan [49] proposed a deep learning model to locate and identify oil spills. The data set was collected through web mining photos from previous oil spills; this data set contained 1,292 images. This research used redgreen-blue (RGB) imagery to minimize the implementation cost. The system used the VGG16 CNN model for classification and mask region-based CNN (R-CNN) model for instance segmentation. Additionally, the scarcity of annotated oil spill and marine object detection data challenged the training of high-performing CNN models. The research indicated that the VGG16 model yields an accuracy of 93% and the mask R-CNN model yields an average precision and recall of 61% and 70%, respectively. Moreover, the results revealed that data analyses and AI can be integrated into upstream and downstream operations in the oil & gas field. However, researchers can use infrared images instead of RGB images as input for the proposed model.

Han et al. [20] Proposed a new method to identify mixed gases based on diverse types of CNN such as ResNet50, ResNet18, GG-16, ResNet34, and VGG-19. A data set comprised of 2970 samples was collected by an array of eight MOX gas sensors. However, several types of CNN were used to classify and compare daffier types of mixed gases. So, after adjusting the parameters, the results showed that the final gas identification rate was 96.67%. The model worked effectively in an environment with large amounts of data. A big challenge was that gas data is time-series data. Also, the insufficient number of samples led to a decrease in average accuracy and a higher error rate. However, different techniques are required to improve accuracy in the case of an insufficient number of samples.

Wang et al. [51] tackled the gas emissions problem using Optical gas imaging (OGI), a well-known, and widely used method for methane leak detection, but labor-intensive, and requires operators' judgment to provide results. A computer vision (CV) based approach with OGI using CNN trained on different methane leak images for automatic detection. A large dataset was built that includes videos of methane leaks from different leakage sources. So, a total of 669.600 frames were recorded. The binary detection accuracy has reached 97% for large gas leaks, and the overall accuracy of all leak sizes has reached 95%. However, for higher accuracy results further work is required to enlarge and develop the dataset to have a diversity of leaks in the real world. Moreover, the exploration of various kinds of model architecture is a critical point.

De Kerf et al. [52] proposed a novel framework to detect the oil that spills inside the port area using a thermal IR camera and an unmuted aerial vehicle (UAV) to decrease the cleaning cost of the oil spill and increase the detection rate. An infrared camera is important to detect oil spills during the nighttime. The dataset containing IR and RGB images was used to train on a CNN, 70% for training, 20% for validation, and 10% for testing. So, 8 different feature extractors and seven different CNN segmentation architectures were used to find the best. However, the proposed model has achieved an accuracy of 89%, which means that the model is efficient in terms of accuracy. However, advanced RGB preprocessing techniques could be implemented, and other camera technologies could be used to improve the accuracy of this model. Table 3 summarizes the intelligent based schemes. Table 3. Summary of intelligent based techniques

Study	Year	Methods	Results
Qu et al. [44]	2010	SVM	Accuracy of 95% for leakage detection, and 200 m precision
			for localization.
Rashid et al. [37]	2014	GMM, SVM, KNN, WSN	Accuracy SVM leak detection 94.73%, size estimation 92.3%.
Mohamed et al. [45]	2015	Static, cascaded, dynamic neural network.	Accuracy over 86% for different error tolerance levels.
Layouni et al. [46]	2017	FFNN	Prediction error is in the range of $\pm 9\%$ of real defect depth.
Chen et al. [40]	2018	SVM, "one-to-one" algorithm	Accuracy (SVM 90%)
Oliveira et al. [34]	2018	Anomaly detection approach and a linear regression ML model.	The proposed system detects pipeline leaks with 3-5 false alarms per month.
Hubert and	2019	RF, GB Trees algorithm	Accuracy of 81% GB Trees
Padovese [29]			
Ahn et al. [38]	2019	SV	100% accuracy for vibration signals while AE signals 80%, 70% for GA and PCA.
Han et al. [50]	2019	CNN as ResNet50, ResNet18, GG-16, ResNet34, and VGG-19	The final gas identification rate was 96.67%.
Wang et al. [51]	2019	CNN	Detection accuracy 97% for large gas leaks, all sizes average accuracy 95%
Akinsete and Oshingbesan [35]	2019	RF, GB, SVM, DT, ANN.	SVM and ANN models have the highest accuracy which is 98%.
Xiao et al. [39]	2019	Relief-F for selecting best features and SVM for the leak intensity.	Accuracy reached 99.4%.
Liu et al. [40]	2019	WSN and SVM	Accuracy reached 98%.
Kampelopoulos et al. [31]	2020	SVM and DT	Recall of (97.07%) by SVM
da Cruz et al. [32]	2020	logistic regression, KNN, SVM-LSVM- RBF, RF, AdaBoost, Xgboost,	Leakage accuracy of 99.6% using RF, rate of false alarms of 0.3%, maximum location error of 4.31% using Xgboost.
Ghorbani and	2020	Deep learning models (mask R-CNN and	R-CNN a recall and average precision of 70% and 61%,
Behzadan [49]		VGG16) to localize and identify oil spills.	respectively. Accuracy of 93% using VGG16.
De Kerf et al. [52]	2020	7 different CNN architectures were used to find the best.	Accuracy of 89%
Wang et al. [36]	2020	LR, DT, RF, AdaBoost.	Accuracy (LR- 70%, and 57% and 66% in recall/sensitivity and balanced).
Spachos et al. [20]	2020	Different architecture models of CNN and screw connections to for gas leakage	Accuracy greater than 93%.
Banjara et al. [43]	2020	SVM, RVM	Accuracy for binary classification SVM- 99.92%, RVM- 98.5%.
Ghorbani and Behzadan [49]	2020	Deep learning models, namely VGG-16 and mask R-CNN	VGG16 model's accuracy of 93% and the Mask R-CNN model's average precision and recall of 61% and 70%, respectively.
Chi et al. [30]	2021	RF, SVM, ANN, k-NN, DT	The RF classifier obtained a classification accuracy 88.33%.
Narkhede and Padovese [29]	2021	CNN, LSTM, and the Early Fusion Model	Accuracy of 96%
Kim et al. [47]	2021	Dynamic model and ML, deep neural network method.	Model accuracy improved by 80% compared to the initial learning model.

## 3.4 Other smart techniques

According to Adegboye et al. [53], the pipeline failure is subject to various reasons. Based on a survey in the study of Bolotina et al. [19], it is mainly caused by corrosion, external factors, human negligence, manufacturing process and installation, as depicted in Figure 4. Out of which the external factors contribute the most, followed by installation issues and corrosion and then come the manufacturing faults while human negligence contributes to 5% only.



Figure 4. Pipeline leakage reasons

Several techniques used for pipeline leakage detection rely on imagery and surveillance of the oil & gas pipelines. For such purpose remote sensing, WSN and unmanned ariel vehicles (UAVs) like drones carrying sensory and/or imagery equipment-based methods are investigated [54].

# 4. DISCUSSION AND CONCLUSION

In this research, more than forty studies on oil & gas pipeline leak detection classified as hardware, software, and intelligent based techniques, have been surveyed. Most of the studies using software and hardware-based techniques used both the IoT and sensors cloud [55-82]. Regarding the intelligent-based studies, they used ML techniques in conjunction with sensors. Therefore, pipeline leak detection can be accomplished using diverse methods and techniques. Though each approach has its own advantages and disadvantages, all leak detection systems were effective for a certain region/type. Nevertheless, most studies had limitations, which should be addressed in future studies. The main limitation observed is that most of the studies did not use a real dataset, but a hypothetical data set created by the researchers, which was derived from sensors and other measures. This resulted in relatively small datasets, providing results that can be improved by increasing the dataset size and features. Furthermore, most studies only focused on leak detection without specifying leak location or size. We noticed that some studies only highlight upstream operation leakage and did not include downstream leakage, such as subsea leakage. Additionally, we did not find any study that applied ensemble methods in which multiple models are combined to produce improved results. In these studies, the researchers used multiple models but trained their systems separately. Subsequently, our findings indicate that the SVM algorithm was the most used and highest performing algorithm, overall. It is also among the most widely used algorithms in other similar areas of research. Although many techniques exhibited sufficient performance, some problems are yet to be resolved. This leaves us in a position to critically evaluate the current solutions and to identify potential research directions that may lead to building models using ensemble-based AI techniques. This research review can be considered as a starting point for researchers and developers to learn about the latest research and techniques in detecting oil & gas pipeline leaks. However, in their future work, multiple researchers intend to create systems that can identify the size and location of the leakage. Additionally, since much of the previous research uses hypothetical data sets derived from sensors and other measures, researchers should consider more realistic datasets from an industry as a fundamental direction to develop expert systems. Finally, this research may aid the researchers in selecting the best techniques and models to build a system to detect oil & gas pipeline leakage.

# REFERENCES

- Mall, A. (2019). Pipeline Incident Statistics Reveal Significant Dangers. Natural Resources Defense Council (NRDC). https://www.nrdc.org/experts/amymall/pipeline-incident-statistics-reveal-significantdangers.
- [2] Spandonidis, C., Theodoropoulos, P., Giannopoulos, F., Galiatsatos, N., Petsa, A. (2022). Evaluation of deep learning approaches for oil & gas pipeline leak detection using wireless sensor networks. Engineering Applications of Artificial Intelligence, 113: 104890. https://doi.org/10.1016/j.engappai.2022.104890
- [3] Muspratt, A. (2019). Introduction to Oil and Gas Industry. Oil & Gas IQ. https://www.oilandgasiq.com/strategymanagement-and-information/articles/oil-gas-industryan-introduction (accessed Nov. 07, 2021).
- [4] "Why are natural gas leaks a problem? Environmental Defense Fund." https://www.edf.org/climate/methanemaps/leaksproblem (accessed Nov. 07, 2021).
- [5] Ahmed, M.I.B., Alotaibi, R.B., Al-Qahtani, R.A., Al-Qahtani, R.S., Al-Hetela, S.S., Al-Matar, K.A., Krishnasamy, G. (2023). Deep learning approach to recyclable products classification: Towards sustainable waste management. Sustainability, 15(14): 11138. https://doi.org/10.3390/su151411138
- [6] "What is Machine Learning? IBM." https://www.ibm.com/cloud/learn/machine-learning

(accessed Oct. 16, 2021).

- [7] "Applying AI in oil and gas." https://www.ey.com/en\_bh/applying-ai-in-oil-and-gas (accessed Nov. 07, 2021).
- [8] Ibrahim, N.M., Gabr, D.G., Rahman, A., Musleh, D., AlKhulaifi, D., AlKharraa, M. (2023). Transfer learning approach to seed taxonomy: A wild plant case study. Big Data and Cognitive Computing, 7(3): 128. https://doi.org/10.3390/bdcc7030128
- [9] Farooqui, M., Rahman, A., Alorefan, R., Alqusser, M., Alzaid, L., Alnajim, S., Althobaiti, A., Ahmed, M.S. (2023). Food classification using deep learning: Presenting a new food segmentation dataset. Mathematical Modelling of Engineering Problems, 10(3): 1017-1024. https://doi.org/10.18280/mmep.100336
- [10] Khan, T.A., Fatima, A., Shahzad, T., Alissa, K., Ghazal, T.M., Al-Sakhnini, M.M., Ahmed, A. (2023). Secure IoMT for disease prediction empowered with transfer learning in healthcare 5.0, the concept and case study. IEEE Access, 11: 39418-39430. https://doi.org/10.1109/ACCESS.2023.3266156
- [11] Somov, A., Baranov, A., Spirjakin, D., Spirjakin, A., Sleptsov, V., Passerone, R. (2013). Deployment and evaluation of a wireless sensor network for methane leak detection. Sensors and Actuators A: Physical, 202: 217-225. https://doi.org/10.1016/j.sna.2012.11.047
- [12] Sun, B., Zhang, P., Guo, L., Fu, M., Fu, M., Cheng, W. (2021). Feasibility study of detecting gas leakage in buried pipeline based on resistivity method. In IOP Conference Series: Earth and Environmental Science, 660(1): 012008. https://doi.org/10.1088/1755-1315/660/1/012008
- [13] Erden, F., Soyer, E.B., Toreyin, B.U., Cetin, A.E. (2010).
   VOC gas leak detection using pyro-electric infrared sensors. In 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, Dallas, TX, USA, pp. 1682-1685. https://doi.org/10.1109/ICASSP.2010.5495500
- [14] Kusriyanto, M., Yulianto, A., Kurniawan, S. (2018).
   Early detection of LPG gas leakage based Wireless Sensor Networking. In MATEC web of Conferences, 154: 01045.
  - https://doi.org/10.1051/MATECCONF/201815401045
- [15] Khan, M.M. (2020). Sensor-based gas leakage detector system. Engineering Proceedings, 2(1): 28. https://doi.org/10.3390/ecsa-7-08278
- [16] Wang, W., Mao, X., Liang, H., Yang, D., Zhang, J., Liu, S. (2021). Experimental research on in-pipe leaks detection of acoustic signature in gas pipelines based on the artificial neural network. Measurement, 183: 109875. https://doi.org/10.1016/j.measurement.2021.109875
- [17] Yan, H.H., Rahayu, Y. (2014). Design and development of gas leakage monitoring system using Arduino and Zigbee. In Proceeding of International Conference on Electrical Engineering, Computer Science and Informatics (EECSI 2014), Yogyakarta, Indonesia, pp. 207–212, https://doi.org/10.11591/eecsi.1.404
- [18] Guo, K., Yang, P., Guo, D.H., Liu, Y. (2019). Gas leakage monitoring with mobile wireless sensor networks. Procedia Computer Science, 154: 430-438. https://doi.org/10.1016/J.PROCS.2019.06.061
- [19] Bolotina, I., Borikov, V., Ivanova, V., Mertins, K., Uchaikin, S. (2018). Application of phased antenna arrays for pipeline leak detection. Journal of Petroleum

Science and Engineering, 161: 497-505. https://doi.org/10.1016/j.petrol.2017.10.059

- [20] Spachos, P. Song, L., Hatzinakos, D. (2014). Gas leak detection and localization system through wireless sensor networks. In 2014 IEEE 11th Consumer Communications and Networking Conference (CCNC), NV. USA. 1130-1131. Las Vegas. pp. https://doi.org/10.1109/CCNC.2014.6940506
- Hou, Q., Ren, L., Jiao, W., Zou, P., Song, G. (2013). An improved negative pressure wave method for natural gas pipeline leak location using FBG based strain sensor and wavelet transform. Mathematical Problems in Engineering, 2013(1): 278794. https://doi.org/10.1155/2013/278794
- [22] Jiang, H., Wang, Y., Deng, J., Xu, M. (2021). A gas pipeline anti-damage early warning monitoring system. In Journal of Physics: Conference Series, 2033(1): 012184. https://doi.org/10.1088/1742-6596/2033/1/012184
- [23] Sharma, M., Tripathi, D., Yadav, N.P., Rastogi, P. (2018). Gas leakage detection and prevention kit provision with IOT. Gas, 5(2): 2227-2230.
- [24] Ralevski, M., Stojkoska, B.R. (2019). IoT based system for detection of gas leakage and house fire in smart kitchen environments. In 2019 27th Telecommunications Forum (TELFOR), Belgrade, Serbia, pp. 2019-2022. https://doi.org/10.1109/TELFOR48224.2019.8971021
- [25] Debnath, S., Ahmed, S., Das, S., Nahid, A.A., Bairagi, A.K. (2020). IoT based low-cost gas leakage, fire, and temperature detection system with call facilities. In 2020 2nd International Conference on Advanced Information and Communication Technology (ICAICT), Dhaka, Bangladesh, pp. 11-16. https://doi.org/10.1109/ICAICT51780.2020.9333530
- [26] Alshammari, B.F., Chughtai, M.T. (2020). IoT gas leakage detector and warning generator. Engineering, Technology & Applied Science Research, 10(4): 6142-6146.
- [27] Swetha, S.V., Swetha, G.N. (2021). LPG/CNG gas leakage detection and prevention using IoT. Int. Res. J. Eng. Technol, 8(4): 3035-3038.
- [28] Babu, T., Nair, R.R., Kishore, S., Vineeth, M. (2024). Enhancing gas leak detection with IoT technology: An innovative approach. Procedia Computer Science, 235: 961-969. https://doi.org/10.1016/j.procs.2024.04.091
- [29] Hubert, P., Padovese, L. (2019). A machine learning approach for underwater gas leakage detection. arXiv preprint arXiv:1904.05661. https://doi.org/10.48550/arXiv.1904.05661
- [30] Chi, Z., Li, Y., Wang, W., Xu, C., Yuan, R. (2021). Detection of water pipeline leakage based on random forest. In Journal of Physics: Conference Series, 1978(1): 012044. https://doi.org/10.1088/1742-6596/1978/1/012044
- [31] Kampelopoulos, D., Papastavrou, G.N., Kousiopoulos, G.P., Karagiorgos, N., Goudos, S.K., Nikolaidis, S. (2020). Machine learning model comparison for leak detection in noisy industrial pipelines. In 2020 9th International Conference on Modern Circuits and Systems Technologies (MOCAST), Bremen, Germany, pp. 1-4. https://doi.org/10.1109/MOCAST49295.2020.9200261

nups://doi.org/10.1109/MOCAS149295.2020.9200261

[32] 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 Technologies and Environmental Policy, 22: 627-638. https://doi.org/10.1007/s10098-019-01805-x

- [33] Narkhede, P., Walambe, R., Mandaokar, S., Chandel, P., Kotecha, K., Ghinea, G. (2021). Gas detection and identification using multimodal artificial intelligence based sensor fusion. Applied System Innovation, 4(1): 3. https://doi.org/10.3390/asi4010003
- [34] Oliveira, E., Fonseca, M., Kappes, D., Medeiros, A., Stefanini, I. (2018). Leak detection system using machine learning techniques. In 6th International Congress on Automation in Mining. Chile, pp. 1-9, 2018.
- [35] Akinsete, O., Oshingbesan, A. (2019). Leak detection in natural gas pipelines using intelligent models. In SPE Nigeria Annual International Conference and Exhibition, Lagos, Nigeria, p. SPE-198738-MS. https://doi.org/10.2118/198738-MS
- [36] Wang, J., Nadarajah, S., Wang, J., Ravikumar, A.P. (2020). A machine learning approach to methane emissions mitigation in the oil and gas industry. In Proc. 34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada, pp. 1-5.
- [37] Rashid, S., Akram, U., Khan, S.A. (2015). WML: Wireless sensor network based machine learning for leakage detection and size estimation. Procedia Computer Science, 63: 171-176. https://doi.org/10.1016/j.procs.2015.08.329
- [38] Ahn, B., Kim, J., Choi, B. (2019). Artificial intelligencebased machine learning considering flow and temperature of the pipeline for leak early detection using acoustic emission. Engineering Fracture Mechanics, 210: 381-392.

https://doi.org/10.1016/j.engfracmech.2018.03.010

- [39] 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. https://doi.org/10.1016/j.measurement.2019.06.050
- [40] Liu, Y., Ma, X., Li, Y., Tie, Y., Zhang, Y., Gao, J. (2019). Water pipeline leakage detection based on machine learning and wireless sensor networks. Sensors, 19(23): 5086. https://doi.org/10.3390/s19235086.
- [41] Chen, Y., Zhao, J., Li, F. (2018). An SVM-based recognition method for safety monitoring signals of oil and gas pipeline. In IOP Conference Series: Materials Science and Engineering, 452(3): 032008. https://doi.org/10.1088/1757-899X/452/3/032008
- [42] Wang, C., Han, F., Zhang, Y., Lu, J. (2020). An SAEbased resampling SVM ensemble learning paradigm for pipeline leakage detection. Neurocomputing, 403: 237-246. https://doi.org/10.1016/j.neucom.2020.04.105
- [43] Banjara, N.K., Sasmal, S., Voggu, S. (2020). Machine learning supported acoustic emission technique for leakage detection in pipelines. International Journal of Pressure Vessels and Piping, 188: 104243. https://doi.org/10.1016/j.ijpvp.2020.104243
- [44] Qu, Z., Feng, H., Zeng, Z., Zhuge, J., Jin, S. (2010). A SVM-based pipeline leakage detection and pre-warning system. Measurement, 43(4): 513-519. https://doi.org/10.1016/j.measurement.2009.12.022
- [45] 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, Rome, Italy, pp. 585-590.

https://doi.org/10.1109/FiCloud.2015.54

- [46] 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. Applied Soft Computing, 52: 247-261. https://doi.org/10.1016/j.asoc.2016.10.040
- [47] Kim, J., Chae, M., Han, J., Park, S., Lee, Y. (2021). The development of leak detection model in subsea gas pipeline using machine learning. Journal of Natural Gas Science and Engineering, 94: 104134. https://doi.org/10.1016/j.jngse.2021.104134
- [48] Song, Y., Li, S. (2021). Gas leak detection in galvanised steel pipe with internal flow noise using convolutional neural network. Process Safety and Environmental Protection,  $146 \cdot$ 736-744. https://doi.org/10.1016/j.psep.2020.11.053
- [49] Ghorbani, Z., Behzadan, A.H. (2020). Identification and instance segmentation of oil spills using deep neural networks. In 5th World Congress on Civil, Structural, and Environmental Engine, Lisbon, Purtogal, pp. 1401-1408. https://doi.org/10.11159/iceptp20.140
- [50] Han, L., Yu, C., Xiao, K., Zhao, X. (2019). A new method of mixed gas identification based on a convolutional neural network for time series classification. Sensors. 19(9): 1960. https://doi.org/10.3390/s19091960
- [51] Wang, J., Tchapmi, L.P., Ravikumar, A.P., McGuire, M., Bell, C.S., Zimmerle, D., Brandt, A.R. (2020). Machine vision for natural gas methane emissions detection using an infrared camera. Applied Energy, 257: 113998. https://doi.org/10.1016/j.apenergy.2019.113998
- [52] De Kerf, T., Gladines, J., Sels, S., Vanlanduit, S. (2020). Oil spill detection using machine learning and infrared images. Remote Sensing, 12(24): 4090. https://doi.org/10.3390/rs12244090
- [53] 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 https://doi.org/10.3390/s19112548
- [54] Boaz, L., Kaijage, S., Sinde, R. (2014). An overview of pipeline leak detection and location systems. In Proceedings of the 2nd Pan African International Conference on Science, Computing and Telecommunications (PACT 2014), Arusha, Tanzania, 133-137. pp. https://doi.org/10.1109/SCAT.2014.7055147

- [55] Alhaidari, F., Rahman, A., Zagrouba, R. (2023). Cloud of Things: architecture, applications and challenges. Journal of Ambient Intelligence and Humanized Computing, 14(5): 5957-5975. https://doi.org/10.1007/s12652-020-02448-3
- [56] Atta-ur-Rahman, Dash, S., Ahmad, M., Iqbal, T. (2021). Mobile cloud computing: a green perspective. In: Udgata, S.K., Sethi, S., Srirama, S.N. (eds) Intelligent Systems. Lecture Notes in Networks and Systems, vol 185. pp. Springer, Singapore. 523-533. https://doi.org/10.1007/978-981-33-6081-5 46
- [57] Dash, S., Sultan, K., Khan, M.A. (2018). Management of resource usage in mobile cloud computing. International Journal of Pure and Applied Mathematics, 119(16): 255-261.
- [58] Dilawari, A., Khan, M.U.G., Al-Otaibi, Y.D., Rehman, Z.U., Rahman, A.U., Nam, Y. (2021). Natural language

description of videos for smart surveillance. Applied Sciences, 11(9): 3730.

https://doi.org/10.3390/app11093730

- [59] Alhaidari, F., Almotiri, S.H., Al Ghamdi, M.A., Khan, M.A., Rehman, A., Abbas, S., Rahman, A.U. (2021). Intelligent software-defined network for cognitive routing optimization using deep extreme learning machine approach. Computers, Materials & Continua, 67(1): 1269-1285. http://doi.org/10.32604/cmc.2021.013303
- [60] Atta-ur-Rahman, Dash, S., Luhach, A.K., Chilamkurti, N., Baek, S., Nam, Y. (2019). A Neuro-fuzzy approach for user behaviour classification and prediction. Journal Computing, of Cloud 17 8. https://doi.org/10.1186/s13677-019-0144-9
- [61] Mahmud, M., Lee, M., Choi, J.Y. (2020). Evolutionarybased image encryption using RNA codons truth table. Optics & Laser Technology, 121: 105818. https://doi.org/10.1016/j.optlastec.2019.105818
- [62] Rahman, A., Qureshi, I., Malik, A., Naseem, M. (2014). A real time adaptive resource allocation scheme for OFDM systems using GRBF-neural networks and fuzzy rule base system. International Arab Journal of Information Technology (IAJIT), 11(6): 593-601.
- [63] Rahman, A.U., Sultan, K., Musleh, D., Aldhafferi, N., Algahtani, A., Mahmud, M. (2018). Robust and fragile medical image watermarking: A joint venture of coding and chaos theories. Journal of Healthcare Engineering, 2018(1): 8137436. https://doi.org/10.1155/2018/8137436.
- [64] Sultan, K., Aldhafferi, N., Alqahtani, A., Mahmud, M. (2018). Reversible and fragile watermarking for medical images. Computational and Mathematical Methods in Medicine, 2018(1): 3461382. https://doi.org/10.1155/2018/3461382
- [65] Qureshi, I.M., Malik, A.N., Naseem, M.T. (2016). QoS and rate enhancement in DVB-S2 using fuzzy rule based system. Journal of Intelligent & Fuzzy Systems, 30(2): 801-810. https://doi.org/10.3233/IFS-151802
- [66] Rahman, A. (2023). GRBF-NN based ambient aware realtime adaptive communication in DVB-S2. Journal of Ambient Intelligence and Humanized Computing, 14(5): 5929-5939. https://doi.org/10.1007/s12652-020-02174-
- [67] Ahmad, M., Qadir, M.A., Rahman, A., Zagrouba, R., Alhaidari, F., Ali, T., Zahid, F. (2023). Enhanced query processing over semantic cache for cloud based relational databases. Journal of Ambient Intelligence and Humanized Computing, 5853-5871. 4: https://doi.org/10.1007/s12652-020-01943-x
- [68] Zaman, G., Mahdin, H., Hussain, K., Abawajy, J., Mostafa, S.A. (2021). An ontological framework for information extraction from diverse scientific sources. IEEE Access. 9: 42111-42124. https://doi.org/10.1109/ACCESS.2021.3063181
- [69] Rahman, A.U., Alqahtani, A., Aldhafferi, N., Nasir, M. U., Khan, M.F., Khan, M.A., Mosavi, A. (2022). Histopathologic oral cancer prediction using oral squamous cell carcinoma biopsy empowered with transfer learning. 22(10): 3833. Sensors, https://doi.org/10.3390/s22103833
- [70] Atta-ur-Rahman, A., Sultan, K., Naseer, I., Majeed, R., Musleh, D., Salam-Gollapalli, M.A., Adnan-Khan, M. (2021). Supervised machine learning-based prediction of

COVID-19. Computers, Materials & Continua, 69(1): 21-34.

- [71] Qureshi, I.M., Malik, A.N., Naseem, M.T. (2014). Dynamic resource allocation in OFDM systems using DE and FRBS. Journal of Intelligent & Fuzzy Systems, 26(4): 2035-2046. https://doi.org/10.3233/IFS-130880
- [72] Rahman, A.U., Dash, S., Luhach, A.K. (2021). Dynamic MODCOD and power allocation in DVB-S2: A hybrid intelligent approach. Telecommunication Systems, 76(1): 49-61. https://doi.org/10.1007/s11235-020-00700-x
- [73] Alhaidari, F.A. (2019). The digital library and the archiving system for educational institutes. Pakistan Journal of Information Management and Libraries, 20: 94-117.
- [74] Rahman, A.U., Abbas, S., Gollapalli, M., Ahmed, R., Aftab, S., Ahmad, M., Mosavi, A. (2022). Rainfall prediction system using machine learning fusion for smart cities. Sensors, 22(9): 3504. https://doi.org/10.3390/s22093504
- [75] Alqarni, A., Rahman, A. (2023). Arabic tweets-based sentiment analysis to investigate the impact of COVID-19 in KSA: A deep learning approach. Big Data and Cognitive Computing, 7(1): 16. https://doi.org/10.3390/bdcc7010016
- [76] Nasir, M.U., Khan, S., Mehmood, S., Khan, M.A., Rahman, A.U., Hwang, S.O. (2022). IoMT-based osteosarcoma cancer detection in histopathology images using transfer learning empowered with blockchain, fog computing, and edge computing. Sensors, 22(14): 5444.

https://doi.org/10.3390/s22145444

- [77] Nasir, M.U., Zubair, M., Ghazal, T.M., Khan, M.F., Ahmad, M., Rahman, A.U., Mansoor, W. (2022). Kidney cancer prediction empowered with blockchain security using transfer learning. Sensors, 22(19): 7483. https://doi.org/10.3390/s22197483
- [78] Gollapalli, M., Musleh, D., Ibrahim, N., Khan, M.A., Abbas, S., Atta, A., Omer, A. (2022). A neuro-fuzzy approach to road traffic congestion prediction. Computers, Materials & Continua, 73(1): 295-310. https://doi.org/10.32604/cmc.2022.027925
- [79] Rahman, A. (2019). Memetic computing based numerical solution to Troesch problem. Journal of Intelligent & Fuzzy Systems, 37(1): 1545-1554. https://doi.org/10.3233/JIFS-18579
- [80] Rahman, A. (2019). Optimum information embedding in digital watermarking. Journal of Intelligent & Fuzzy Systems, 37(1): 553-564. https://doi.org/10.3233/JIFS-162405
- [81] Atta-ur-Rahman. (2018). Efficient decision based spectrum mobility scheme for cognitive radio based V2V communication system. Journal of Communications, 13(9): 498-504.
- [82] Talha, M., Sarfraz, M., Rahman, A., Ghauri, S.A., Mohammad, R.M., Krishnasamy, G., Alkharraa, M. (2023). Voting-based deep convolutional neural networks (VB-DCNNs) for M-QAM and M-PSK signals classification. Electronics, 12(8): 1913. https://doi.org/10.3390/electronics12081913