Journal of Pipeline Science and Engineering 1 (2021) 436–451

Contents lists available at ScienceDirect

Journal of Pipeline Science and Engineering

journal homepage: http://www.keaipublishing.com/en/journals/journal-of-pipeline-science-and-engineering/

Current technologies and the applications of data analytics for crude oil leak detection in surface pipelines

Francis Idachaba*, Minou Rabiei

Department of Petroleum Engineering, University of North Dakota, Grand Forks, North Dakota, United States

ARTICLE INFO

Keywords:
Leak detection
Sensor data analysis
Real time leak detection
Surface pipelines

ABSTRACT

Pipeline pressure monitoring has been the traditional and most popular leak detection approach, however, the delays with leak detection and localization coupled with the large number of false alarms led to the development of other sensor-based detection technologies. The Real Time Transient Model (RTTM) currently has the best performance metric, but it requires collection and analysis of large data volume which, in turn, has an impact in the detection speed. Several data mining (DM) methods have been used for leak detection algorithm development with each having its own advantages and shortcomings. Mathematical modelling is used for the generation of simulation data and this data is used to train the leak detection and localization models. Mathematical models and simulation software have also been shown to provide comparable results with experimental data with very high levels of accuracy. While the ANN and SVM require a large training dataset for development of accurate models, mathematical modelling has been shown to be able to generate the required datasets to justify the application of data analytics for the development of model-based leak detection systems for petroleum pipelines. This paper presents a review of key leak detection strategies for oil and gas pipelines, with a specific focus on crude oil applications, and presents the opportunities for the use of data analytics tools and mathematical modelling for the development of a robust real time leak detection and localization system for surface pipelines. Several case studies are also presented.

1. Introduction

Wherever there are pipelines, there will be a high probability for the occurrence of leaks. This assertion has been validated over the years as shown in Figs. 1 and 2. Fig. 1 shows the pipeline network of the United States while Fig. 2 shows the locations of pipeline leaks in the United States. The cause of these leaks ranges from man-made activities such as vandalization or inadequate protection of the pipeline against the elements during installation to the corrosion induced failures of the lines caused either by ageing of the lines or operational reasons such as weld failures.

The total length of the global trunk/transmission pipeline network is estimated to be 2,034,065.0 km according to GlobalData reports. Crude oil pipelines make up over 379,000 km while petroleum products pipelines constitute over 267,000 km. Natural gas pipeline have the highest number with almost 1,300,000 km and natural gas liquids constitute over 92,000 km (GlobalData, 2019).

From the reports, North America has the highest oil and gas pipelines length of 834,152.5 km (with start years up to 2023), of which, crude oil pipelines constitute 154,200.9 km, petroleum products pipelines constitute 103,106.3 km, natural gas pipelines constitute 495,555.3 km and NGL pipelines constitute 81,290.0 km. The region's share in the global transmission pipeline length is 41.0%.

From the data shown in Fig. 2, in spite of the advantages provided by pipelines for rapid crude oil transportation, the risks associated with pipeline leaks present a significant challenge to both the companies and the environment. These challenges justify the need for continuous research to develop faster and more accurate leak detection and oil spill containment systems to minimize the impact of pipeline oil spills both on the environment and the operator finances. This paper reviews the major technologies used currently for the pipeline leak detection and the use and development of the data analytics for leak detection based on collection and analysis of large volume of data.

2. Regulations

In view of the benefits of the use of petroleum pipelines and the impact of these spills on the environment, several countries developed specific guidelines and regulations to manage the operation of these pipelines to ensure both the safety of the pipelines and the also that of the environment where these pipelines traverse. These regulations and their countries of origin are as follows.

* Corresponding author.
E-mail addresses: francis.idachaba@und.edu (F. Idachaba), minou.rabiei@und.edu (M. Rabiei).

https://doi.org/10.1016/j.jpse.2021.10.001
Received 1 September 2021; Received in revised form 13 October 2021; Accepted 15 October 2021
2667-1433/

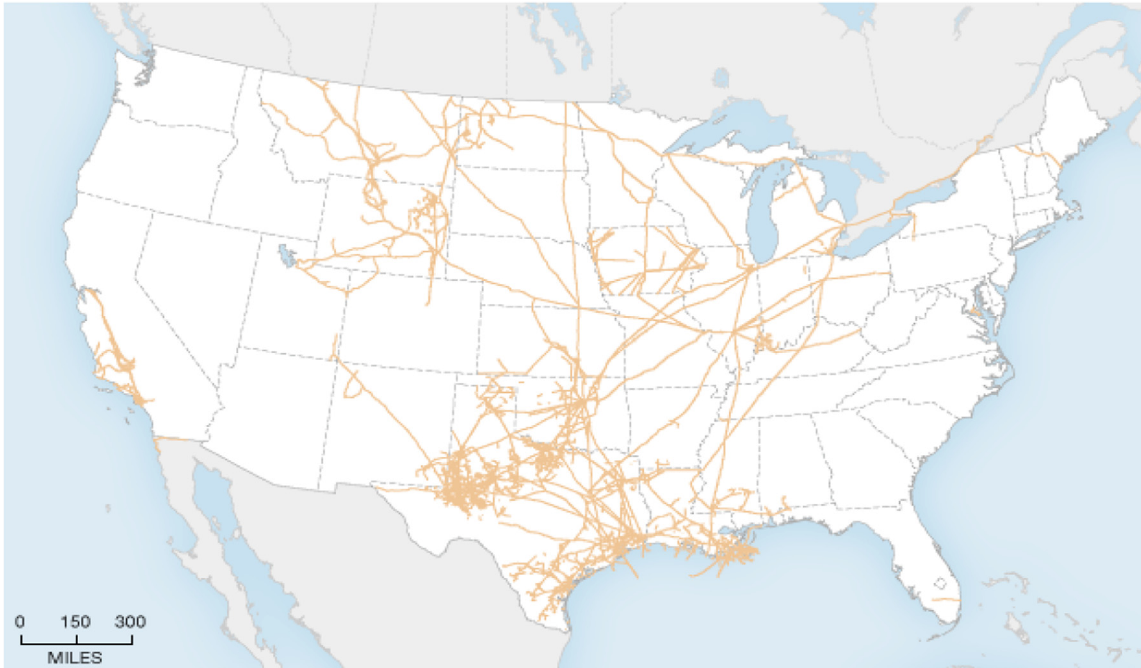

Fig. 1. Pipeline network in the USA (Matthew, 2013).

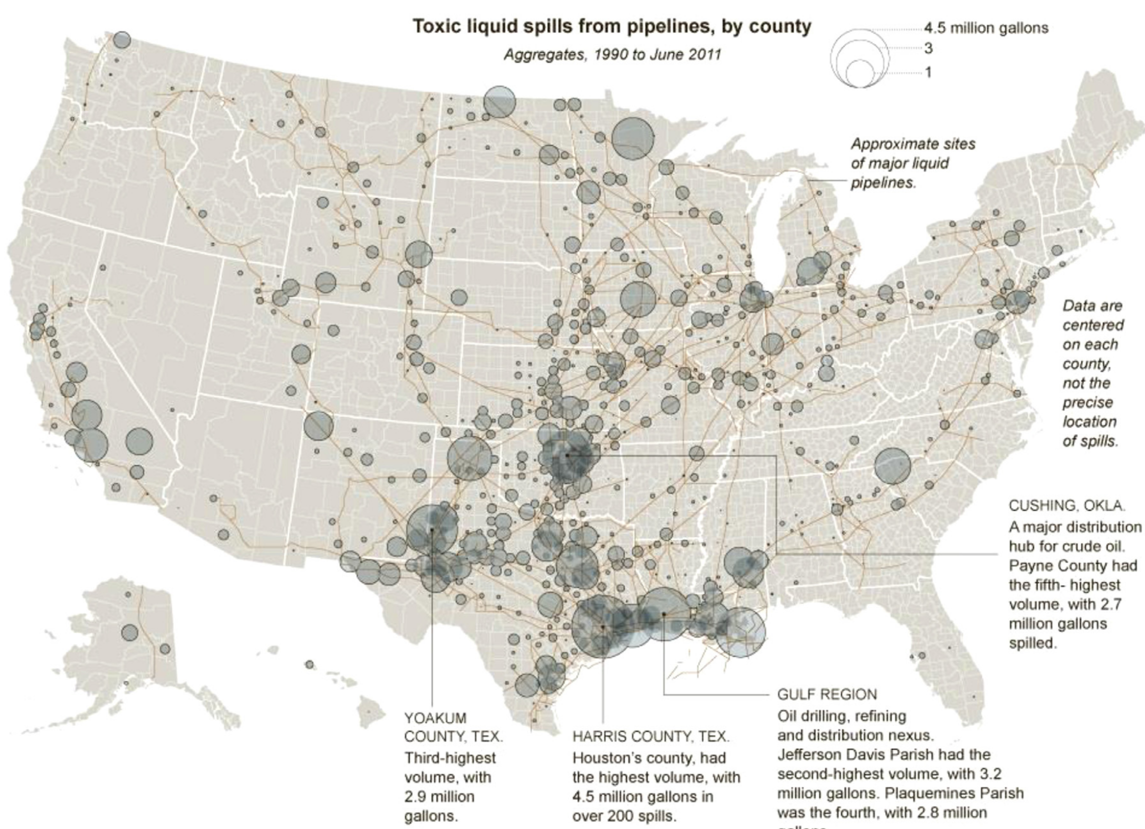

Fig. 2. USA pipeline leak locations (Matthew, 2013).

1) Germany: Germany has the TRFL, the Technical Rule for Pipelines.
2) United States of America: The US has three API standards which guide different parameters in the deployment and operation of petroleum pipelines. These regulations are:
   (1) API 1130, which deals with computational pipeline monitoring for liquids.
   (2) API 1149, which deals with variable uncertainties in pipelines and their effects on leak detection performance.
   (3) The former API 1155, which contains performance criteria for leak detection systems, which has since been replaced by API 1130.

(4) American 49 CFR 195, which regulates the transport of hazardous liquids via pipeline.
3) Canada: CSA Z662, focuses on oil and gas pipelines.

The API 1130 has four primary requirements for a Leak Detection System (LDS) as presented by Lunger and Karami (2019).

1) **Accuracy:** the LDS should be able to calculate leak size and leak location accurately. This is quantified as the maximum distance allowed between estimated location and actual location, as well as the maximum variation allowed between estimated and actual leak size.
2) **Reliability:** the LDS should correctly display any real alarms, but also not report any false alarms. This should be quantified in the number of false alarms acceptable.
3) **Robustness:** the LDS should be able to operate in non-ideal environments, such as when sensor input equipment fails to provide data. This should be quantified in% availability.
4) **Sensitivity:** the LDS should be able to detect small leaks and detect them quickly. This is quantified in absolute flow rate terms because a relative change can be misleading.

The performance criteria of the API 1155 which was replaced by the API 1130 is shown in Table 1.

3. Leak detection classification

Leak detection is the process of detecting the onset of leaks from pipelines. There are two broad classifications of the leak detection systems, these include continuous and non-continuous systems. The non-continuous systems comprise of : inspection by helicopter, smart pigging, tracking dogs, and right of way (ROW) monitoring and patrol (Baroudi et al., 2019).

This approach is usually triggered by a drop-in pressure indicative of a leak event or scheduled routine surveillance. The continuous method can be further classified into internal and external based systems depending on the location of the sensors. The internal systems include the use of: pressure point analysis, mass balance method, statistical systems, RTTM based systems, and extended RTTM.

The external systems include: fibre optic cables, acoustic systems, video monitoring and semi-permeable sensor hoses.

Often, the continuous and non-continuous systems are used together (Romero-Tapia and Fuente, 2018).

The following section presents an overview of the different pipeline leak detection methods.

3.1. Non-continuous systems

3.1.1. Inspection by helicopter

This approach involves the use of helicopters for overfly inspection. The helicopter flies along the pipeline right of way, looking to detect any outflowing oil or gas spills. Special cameras are mounted on the helicopters as they fly over these lines to capture the images emanating from the lines. Three common methods when detecting leaks by helicopter include detection using laser, infrared cameras and “leak sniffers”. When using lasers for leak detection, a laser is set to the absorption wavelength of the medium to be detected. When the laser hits the medium, a part of the laser energy is absorbed. The amount of energy absorbed from the laser is measured to arrive at the amount of leaked medium (Fiedler, 2016). The leak sniffers are used to acquire samples from the leak point for analysis. To accomplish these tasks, the helicopters must fly at very low altitudes and must go through the gas cloud. Some disadvantages of this system include the high cost of helicopter rentals and the inability of flight to take place during poor weather conditions. There is the possibility of the leaked gas being carried away from the leak point by high winds. This system is costly to deploy and is not suitable for real time leak detection. The use of Unmanned Aerial Vehicles (UAVs) has replaced the helicopters in many operations as they are cheaper to deploy and can carry several specialized cameras for pipeline monitoring. The UAVs range from very small portable quadcopters which have low range and can only carry video recording equipment to bigger sized UAVs with capacity for large cameras and on-board gas processing equipment. This has greatly reduced the costs associated with pipeline inspections by helicopter and UAVs (Idachaba, 2016).

Table 1
API 1155 Performance criteria (Fiedler, 2016).

Label	Description
Sensitivity	Minimum detectable leak rate; detection time
Reliability	Avoid false alarms; reliable detect leaks
Accuracy	Accurate localization of leaks
Robustness	Detect failing sensors; fall-back strategies in the event of sensor failure

3.1.2. Inspection by PIGs

Pipeline Inspection Gauges (PIGs) are tools launched into the pipeline carrying specialized inspection equipment for monitoring and reporting the state of the pipeline internal surface. PIGs are also used for cleaning the pipelines, separating product batches, as well as gauging pipeline condition. They enable operators determine information on corrosion, cracks, wall thickness as well as existing leaks in pipelines. The process of pigging includes the insertion of the pig into the pipeline using a pig launcher. The pig advances through the pipeline, propelled by the medium and gathers data along the way. A receiver is used to guide the pig out of the pipeline to subsequently analyse the collected data. Various techniques are used to collect pipeline information using smart pigs; two of the most common are the magnetic flux leakage method and the ultrasonic principle. For the magnetic flux leakage method, the pipe is magnetized by a strong permanent magnet and as the pig passes through the pipe, it detects and monitors any changes in the magnetic flux of that section of the pipe wall. Such changes are indicative of an abnormality or corrosion on the pipe. For the ultrasonic principle, the pig transmits ultrasonic pulses into the pipeline wall and receives their reflected signals. The signals are reflected by both the inner and outer pipe walls and based on the running speed of the pig; the thickness of the pipe wall can be derived. Variations in the pipe wall thickness is interpreted to be caused by a leak, a damage or a corrosion induced failure. A key requirement for the deployment of pigs is that the entire pipe section through which the pig is to be deployed must be piggable. The pipe bends and the valves along that pipe section must allow the free passage of the pig. There must also be additional installation of pig launchers and receivers installed on the lines and the wax build up in the line must not be able to impede the movement of the pig as if goes through the pipe. The speed of the pig must also be kept at between 3 and 15 feet per second to ensure accuracy of the results.

3.1.3. Inspection by tracking dogs

Tracking dogs are specially trained dogs that can detect and distinguish specific gas or hydrocarbon. These dogs are trained on these compounds and are released on the pipeline right of way to sniff out the presence of these spills if it exists. The limitation of this system includes the fact that the dogs can only be deployed on onshore facilities and on short pipelines or segments of pipeline, it is also difficult to certify a tracking dog as a leak detection system within the framework of API or TRFL.

3.1.4. Right of way monitoring and patrol

This system includes the use of local contractors or members of the pipeline host communities to patrol and monitor sections of the pipelines passing through their communities. This approach is suitable for pipelines in locations with a high possibility of vandalization and restive youths. The operators engage the youths as a form of job creation and pays them to monitor the pipelines and report any criminal activity along the pipeline right of way. These contractors are also able to report the commence of a leak as soon as it is detected and as such help to minimize the environmental impact of such leaks. This system is not a real time system and there is no mean of ensuring that the contractors are patrolling the pipeline right of way, however the operators have some form of assurances that they have persons in the communities who will provide them with accurate feedback about their pipes thus reducing the HSE exposure of the operator staff.

3.2. Continuous systems

The continuous system includes the use of external devices for pipeline data acquisition and internal devices which rely on mathematical models and the data acquired from the sensors for detecting and localizing the pipeline leaks. These external systems are as follows.

3.2.1. Optic fibre cables

Optic fibre cables are installed beside the pipeline to utilize their acoustic and temperature sensing characteristics for the continuous external monitoring of pipelines for leaks. The sensor system sends light pulses along the optic fibre and any external vibration due to leaks or intruder activities interferes with the light pulses resulting in reflections which are picked up by the system and interpreted to determine the location of such activities. Oil leaks rely in the vibration and the acoustic signature of the leaks while gas leaks are detected through the temperature differential around the leak point. A temperature profile can be made, and it is then possible to detect the characteristic change in temperature that occurs at the leak site. The system provides accurate leak localization but must follow specific installation guidelines such as the installation of the fibre on top of the pipeline for gas detection since gas rises. Other challenges associated with this system is the high CAPEX cost associated with both the procurement and the installation of the cables as it must cover the entire pipeline right of way to be able to monitor the entire pipeline network.

3.2.2. Acoustic sensors

The escape of oil and gas from a pipeline leak point is accompanied by acoustic signals. Special acoustic sensors are installed outside of the pipeline to detect leaks by measuring the noise levels at multiple sites along the pipeline. This information is used to create a noise profile of the pipeline. The system works by comparing the received signals with the baseline noise profile of the pipeline. These deviations are translated to leak alarms when they exceed specific thresholds. The sensors can be mounted directly on the pipelines or coupled to the pipe wall using steel rods. The longer the pipeline network, the greater the number of acoustic sensors needed for the monitoring. Other challenges associated with this system includes the exposure of the sensors to vandalization, the inability of the system to detect small leaks whose acoustic signals differ slightly from the background noise and the possibility of several false alarms as external noise sources cannot be controlled or eliminated from the pipe surrounding.

3.2.3. Video monitoring

The use of video monitoring for pipelines is restricted to short distances. The video system relies on the use of infrared sensitive filters which can detect the leaks and report them as a smoke image on the video display screen. The difference in the thermal conductivity of wet ground from dry ground also makes it possible for the use of infrared cameras for detecting liquid leaks. This system is limited to critical areas such as on company premises or for high consequence areas.

3.2.4. Semi permeable sensor hoses

Sensor hoses are semi-permeable hoses installed along the pipeline to detect the occurrence of leaks. When a leak occurs, the hydrocarbon discharge from the pipe enters the hose and this discharge combines with a test gas which is injected at specific time intervals. The hose system has an analysing unit at the end which then tests the hose contents for the presence of hydrocarbons. The run time of the test gas injected at the inlet and the time for the gas to get to the analysing unit is used to derive the leak point. The sensitivity of the analysing unit is high enabling it to detect very small volumes of leak but the material-specific properties of the hose, limits the use of sensor hoses to short pipelines (Fiedler, 2016).

4. Leak detection strategies

The most predominant leak detection strategy is the pipeline pressure profile. The occurrence of a leak on a pipeline causes a sudden decrease in the pressure which is followed by a partial recovery to its original value. This pressure pulse travels upstream and downstream through the pipeline as a wave. Leak detection, hydraulic transient wave propagation velocity and leak location are the basic problems to deal when an automatic supervision of a pipeline is to be implemented (Silva et al., 1996). Operators monitor the pipeline pressure profile using the pressure sensors installed at different points on the pipeline. These values are compared with the pressure profile generated from the modelling of the fluid flow in the pipeline using the Darcey Weisbach equation, the pipeline friction factor, and the flow parameters of the fluid in the line. Comparing these two values enables operators to detect anomalies which are indicative of the presence of pipeline leaks. This method is, however, not capable of providing leak localization and as such the identification of a possible leak scenario would require further evaluation and verification before a pipeline shutdown can be considered. The complex nature and the length of the pipeline networks means that each line will comprise of multiple compressors, pumps and these components introduce delays and anomalies in the pressure profile thus making it unreliable for real time leak detection. This challenge led to the development of other leak detection strategies which would enable operators to determine the commencement of leaks with greater accuracy and be able to identify the location of these leaks. The leak detection strategies can also be classified as either software or hardware based (Cui et al., 2018).

The hardware-based approach comprises of the installation of physical sensors on the pipeline, or the pipeline ROW and these sensors detect the variation in pipeline pressure, as in the case of pressure sensors, or detect the presence of leak, as in the case of the fibre optic cables. The optic fibre approach utilizes an optic fibre cable buried beside the pipeline ROW and can detect the leaks and other third-party intrusion by picking the vibration signature of these activities. The data is transmitted to the office location through the optic fibre cable and interpreted to determine the leak point. While this is accurate and fast, the systems are mostly proprietary, require high capital investment to install and the vandalization of any section of the cable can shut down the entire detection system. This technology is known as the Distributed Acoustic System (DAS). The Distributed Temperature Sensing (DTS) feature of the optic fibre cable is used to detect the presence of gas leaks utilizing the Joule Thomson effect which creates a temperature differential (lower temperature) around the leak point as the gas exits the lines during the leaks. This method is only suitable for gas leaks. The performance of the leak detection systems is summarised in Table 2.

From Table 2, it is seen that the pressure point analysis and mass balance methods produce very high rate of false alarm. This is often due to the complex nature of the pipeline comprising of pumps compressors and pipe bends. These all introduce delays in the flow resulting in false leak signatures or very long delay time to detect the leaks.

The real time systems, on the other hand, have very high detection speeds with slight and average rate of false alarms. These systems rely on the real time data acquired from the field sensors on the pipeline and are not subject to the delays introduced by the length of the pipeline or its associated components such as valves.

Fig. 3 shows the general classification of leak detection systems. While some methods, such as the visual methods, are suitable for short pipe lengths, other methods e.g., the external and electromagnetic meth-

Table 2
Performance characteristic of LDS systems (Fiedler, 2016).

Method	Typical Minimum Detectable leak rate	Time to detect leak		Detectable types of leaks	False Alarm frequency
		Liquids	Gases		
Pressure Point Analysis	>5%	Short	Long	Spontaneous leaks	High
Mass Balance Method	>1%	Long	Very Long	Spontaneous and Creeping leaks	High
Statistical Methods	>0.5%	Long	Very Long	Spontaneous and Creeping leaks	Slight
Real Time Transient Model (RTTM)	>1%	Short	Short	Spontaneous and Creeping leaks	Average
E-RTTM	>0.5%	Very short	short	Spontaneous and Creeping leaks	Slight

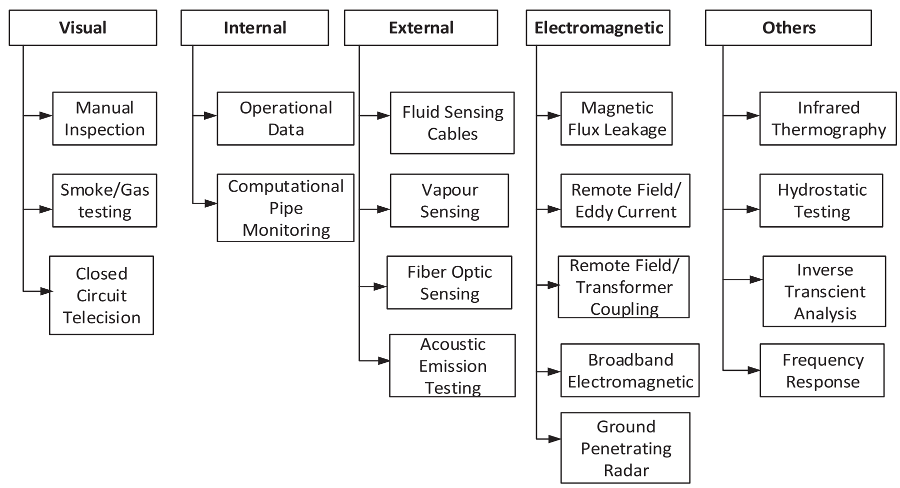

Fig. 3. Classification of leak detection system (Baroudi et al., 2019).

ods require specialized equipment which are costly to acquire and install and may not be able to provide real time pipeline monitoring.

The internal method shown in Fig. 4, on the other hand, relies on operational data and pipeline models to determine the leak points. It requires minimal physical installations and has a rapid detection (real-time) time when compared with the other methods.

Internal systems rely on field sensors to monitor the operational and hydraulic conditions of the pipeline, e.g., measurements of the flow, pressure, and temperature. These real time parameters are compared with the normal working parameters of the pipeline which are determined either manually by pipeline controllers or based on sophisticated algorithms and hydraulic models (Baroudi et al., 2019). A difference between the measured and predicted operational parameters indicates a leak. The remote field sensors installed on the pipeline monitor the lines continuously and send the data to a centralized monitoring station, where the data undergoes filtering, signal processing and is passed on to the leak detection and localization algorithms to both detect the presence of the leaks and identify its location.

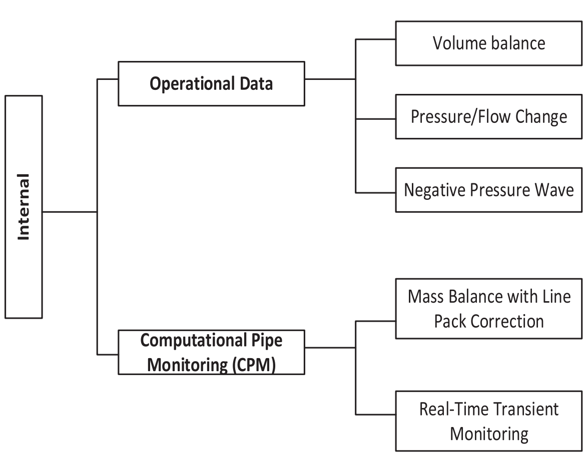

Fig. 4. Classification of internal leak detection systems (Baroudi et al., 2019).

The data acquisition methods from the sensors include the following four components.

1) **Volume Balance:** This comprises of the volume differential between the incoming and outgoing volumes. Volume balance can detect catastrophic failures; however, its usage is rare due to its limited performance.
2) **Rate of Pressure and Flow Change**: Pipeline leaks are characterized by sudden change in pressure, and this can be used to indicate the presence of leaks. However, sudden pressure variations can also be due to transient conditions of the pipeline. Filtering techniques and suitable algorithms are used to differentiate between leaks and operations induced pressure changes. Pressure waves also dampen out as they traverse a longer length and thus additional pressure sensors need to be installed along the pipelines.
3) **Negative Pressure Wave (NPW):** Negative pressure waves are created by sudden leaks, and they propagate in both directions from the leak. A critical challenge of this system is that it cannot differentiate between leaks and normal operations, and this results in false alarms.
4) **Computational Pipeline Monitoring (CPM)**: This method detects anomalies in pipeline operating parameters, and this is accomplished using the Mass Balance with line pack approach and the RTTM method (Baroudi et al., 2019).

The mass balance with line pack correction monitors the pipeline using sensors such as pressure, temperature, densitometer, and other parameters. These sensors are installed at multiple locations on the pipeline between the inlet and outlet flow meters and the parameters of the topography of each pipeline is factored into the model for such locations. The changes measured by various sensors are adjusted in the mass balance to account for transient flows, anticipated fluid changes, and other flow conditions. In the RTTM, the real time field data are compared with the parameters generated by the simulation model and any difference or discrepancy is interpreted as a leak. It has a very high sensitivity and response time as very small leaks can be detected. However, the RTTM system requires extensive training and skilled workers to both operate and maintain it.

Current trends in pipeline leak detection and localizations methods further identifies two main categories for pipeline lead detection under the internal method using real time transient modelling and pipeline data. These are the signal-based methods and the model-based methods. These methods are based on the steady state models of the pipelines, they can predict unknown parameters and determine the occurrence and localization of leaks more accurately. The pipeline-model-based method predicts pressure distribution along the pipeline and locates the leak through the pressure and flow rate signal on both ends of the pipeline. The most popular and widely used pipeline model-based method includes the pressure gradient (PG) method and the average friction coefficient (AFC) method (Cui et al., 2018). The PG method ignores the influences of friction coefficient, temperature, pipe diameter and other factors on the pressure distribution along the pipeline and considers the pressure to be linear. The AFC method, however, assumes the friction coefficient and pipe diameter to be constant (Cui et al., 2018).

The PG method was introduced by Seiders et al. (1979). Two pressure sensors were installed upstream and downstream on a pipe section to measure the PG. The intersection of the two lines identified the leak location. The challenge with this method was the fact that the pressure gradient is not usually a straight line, and this led to the questions about the accuracy of the results. Zhang et al. (1993) revised the PG method and successfully applied it to the oil pipelines of Shell Company. Several other researchers such as Begovich (2010, 2012) and Physiker (2003) further improved the location accuracy. Physiker (2003) was able to estimate the flow rate at both ends of the pipeline. A key limitation with these approaches is the fact that the methods were not applicable to gas pipelines. This is because while the methods were suitable for high pressure liquid lines, gas lines were not liquid and are characterised by negative pressure (Cui et al., 2018). Table 3 shows a summary of all external leak detection methods.

5. Pipeline leak modelling

The occurrence of a leak in pipeline causes a sudden decrease in the pressure which is followed by a partial recovery to its original value. This pressure pulse travels upstream and downstream through the pipeline as a wave. Leak detection, hydraulic transient wave propagation velocity and leak location are the component of the system to be considered for an automatic supervision of a (Silva et al., 1996). The analysis of fluid flow in pipelines for leak detection research requires actual leak data to train the leak detection model. However, such data is very scarce and difficult to obtain from oil and gas fields and operators. This provides the justification for the use of hypothetical data generated by the transmission pipeline model (Sukarno et al., 2007). This generated data is utilized for training the leak detection and localization models by applying the various physical configurations of the pipeline and the oil properties from which an accurate pipeline can be simulated. The data is then used in the development of a realistic and accurate leak detection and localization model.

The transient pipeline flow model provides the foundation for pipeline simulation and modelling. The basic equations governing this model include the continuity, the momentum, and the energy equation and the equation of state.

The continuity equation focuses on the conservation of mass principle. It requires that the difference in mass flow into and out of any section of the pipeline is equal to the rate of change of mass within the section. This can be expressed mathematically as shown:

$$\frac{d\rho}{dt} + \rho\frac{dv}{ds} = 0 \tag{1}$$

Where ρ is density, t is time, v is flow velocity and s is pipeline location coordinates.

The conservation of momentum equation is represented as shown:

$$\frac{dv}{dt} + \frac{1}{\rho}\frac{dp}{ds} + fs = 0 \tag{2}$$

Where p is pressure, and fs is pipeline friction.

The conservation of energy principle is represented in as shown:

$$\frac{dh}{dt} - \frac{1}{\rho}\frac{dp}{dt} - I_L = 0 \tag{3}$$

Where h is enthalpy, and I_L is specific loss performance L.

Eqs. (1) to (3) are the basic equations for one dimensional pipe flow analysis and are present in one form or the order in all transient pipe models. These equations are used in developing the required pipeline model for the specific fluid being transported.

The pipeline network and the governing equations enable the simulation of oil flow in the pipeline and predicts the pressure distribution along the pipe with or without leak under specific flow conditions.

The simulation of a leak is accomplished by introducing a branch pipe of a given diameter on the main pipeline. This branch pipe can be located at any point on the main line with the leakage rate made variable. The variable leakage rate enables the study of different leak types on the main pipeline. This model is represented in Fig. 5 schematically. In this Figure, D_1 is the distance between leak point and upstream pressure sensor and D_2 is the distance between leak point and downstream pressure Sensor.

The flow pattern in the pipeline is represented by the Reynolds number as:

$$Re = \frac{\rho v D}{\mu} \tag{4}$$

Where μ is the fluid viscosity, and D is pipe diameter.

If $Re < 2{,}000$, the flow pattern is called laminar, whereas, if $Re > 4{,}000$, the flow pattern is called turbulent. If $2{,}000 < Re < 4{,}000$, the flow is called transition. This parameter will be utilized in the modelling of the pipeline.

This pipeline model can produce sufficient data for different leak conditions and the data will be used for training the leak detection model. Several machine learning models can utilize the generated data for the development of the leak detection system. The model shown in Fig. 5 for a pipe section can be extended to cover the entire pipeline network and be extended to enable the detection of multiple pipeline leaks from pipelines. The installation for entire pipeline network is shown schematically in Fig. 6.

The pipeline section in Fig. 6 comprises of sensors installed at specific points. The goal of the simulation of this pipeline network is to determine the minimum distance for installation of sensors which can provide end to end leak detection from multiple leak sources.

Once the data has been generated from the pipeline transmission model, the data machine learning pipeline, shown in Fig. 7, will be utilized to develop the leak detection model which will be deployed for

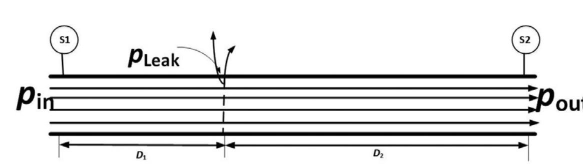

Fig. 5. Section of the pipeline with the leak at distance D_1 from the inflow section.

Table 3
A summary of all the external leak detection methods (Adegboye et al., 2019).

Methods	Principle of Operation	Strengths	Weaknesses
Acoustic Emissions	Detects leaks by picking up intrinsic signals escaping from a perforated pipeline	Easy to install and suitable for early detection, portable and cost effective	Sensitive to random and environmental noise, prone to false alarms and not suitable for small leaks
fibre optic sensing	Detects leaks through the identification of temperature changes in the optical property of the cable induced by the presence of a leak	Insensitive to electromagnetic noise and the optical fibre can act both as a sensor and a data transmission medium	The cost of implementation is high, not durable, and not applicable for pipelines protected by cathodic protection systems
vapour sampling	Utilizes hydrocarbon vapour diffused into the sensor tube to detect trace concentrations of specific hydrocarbon compounds	Suitable for detecting small concentrations of diffused gases	Time taken to detect a leak is long, not effective for subsea pipelines
Infrared Thermography	Detect leaks using infrared image techniques for detecting temperature variations in the pipeline environment	Highly efficient power for transforming detected objects into visual images, easy to use and fast response time	Quantifying leak orifices smaller than 1.0 mm using IRT-based systems is difficult
Ground Penetration Radar	Utilise electromagnetic waves transmitted into the monitoring object by means of moving an antenna along a surface	Timely detection of leakage in underground pipelines, reliable and leak information is comprehensive	GPR signals can easily be distorted in a clay soil environment, costly and require highly skilled operator
Florescence	Proportionality between the amount of fluid discharged and rate of light emitted at a different wavelength	High spatial coverage, quick and easy scanning for leaks	Medium to be detected must be naturally fluorescent
Electromechanical Impedance	Utilize mechanical impedance changes deduced by the incident of pipeline defect	A single piezoelectric transducer can serve as both sensor and actuator	It is only applicable for metal pipelines, operational limitations in high temperature environments
Capacitive Sensing	Measuring changes in the dielectric constant of the medium surrounding the sensor	It can be employed for detection in non-metallic targets	Requires direct contact with the leaking medium
Spectral Scanners	Comparing spectral signature against normal background	Capable of identification of oil type (light/crude) and thickness of the oil slick	The amount of data generated by a spectral scanner is large which limited its ability to operate in nearly real-time
Lidar Systems	Employed pulsed laser as the illumination source for methane detection	Able to detect leaks in the absence of temperature variation between the gas and the surroundings	High cost of execution and false alarm rate
Electromagnetic Reflection	Measure emitted energy at different wavelengths	It can indicate leak location	It can be affected by severe weather

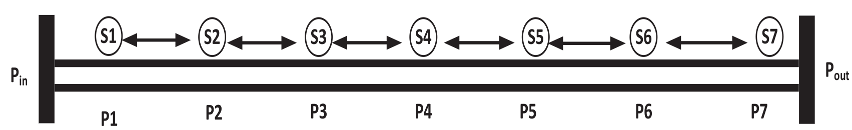

Fig. 6. Pipeline network modelling for leak detection.

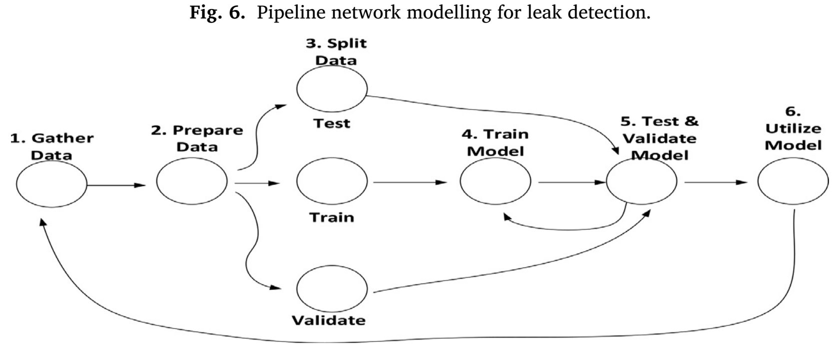

Fig. 7. Machine learning pipeline (Akerkar, 2019).

the detection of leaks from the pipeline. The data will also be used in developing the leak localization module.

Once the model has been finalized, tested, and validated, it is deployed to monitor the pipeline and develop a database of the sensor readings from the pipelines. The model will be able to learn from the data and optimize itself by increasing both its accuracy and speed of detection. The continuous data acquisition leads to the development of the RTTM approach for leak detection.

6. Mathematical modelling of oil and gas pipelines

Pipeline leaks can be accurately modelled using mathematical modelling of the flow in a gas or oil pipeline. The leaks can be accurately predicted by analysing the variations of flow variables such as pressures and flow rates at the ends of a pipeline following the occurrence of a leak. The model can also be used to generate test data for validating model-based leak detection and location methods (Sun, 2012). Utilizing the negative pressure generated by the leak, the following analysis reported by Sun (2012) shows the models for the oil and gas lines with and without leaks are shown in Fig. 8.

Assuming a uniform cross-sectional area along the pipeline, the flow in a gas pipeline without leaks is represented as:

$$\frac{{p_Q}^2 - {p_Z}^2}{G^2RT} - 2\ln\frac{p_Q}{p_Z} = \frac{\lambda L}{D} \tag{5}$$

Where λ is the average friction coefficient along the length of the pipeline, G is the mass flux through the pipeline, p_Q is the inlet pres-

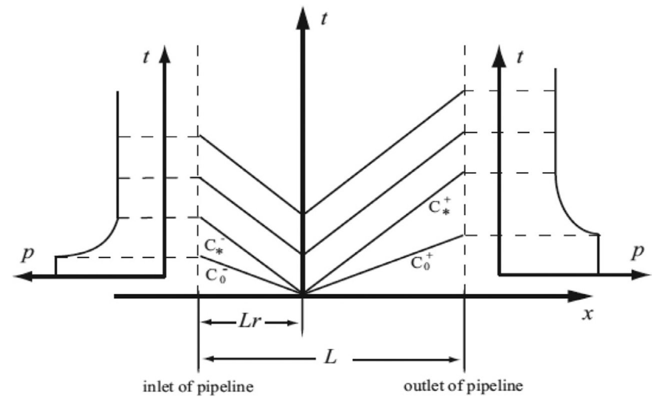

Fig. 8. Generation and propagation of an expanded wave from the leakage point (Sun, 2012).

sure, p_Z is the outlet pressure, R is the universal gas constant and T is the average temperature along the pipeline.

The presence of a leak introduces a negative pressure wave which travels to both ends of the line. Analysing the wave in the C_*^- path, induced by the leak, the equation for the transient flow under leak conditions is given as:

$$\frac{p_Q{}^2 - p_L{}^2}{G_Q{}^2 RT} - 2\ln\frac{p_Q}{p_L} = \frac{\lambda L_r}{D} \tag{6}$$

Where p_L is the pressure at the leak point.G_Q is the upstream mass flow of the leak point. L_r is the distance between the inlet and the leak point. Computing for the flow along the C_*^+ path, the expression is shown as:

$$\frac{p_L{}^2 - p_Z{}^2}{G_Z{}^2 RT} - 2\ln\frac{p_L}{p_Z} = \frac{\lambda(L - L_r)}{D} \tag{7}$$

Where G_z is the mass flow rate, downstream of the leak point.

The leak rate is determined from the expression in Eq. (8) given that the cross-sectional area of the pipeline remains constant.

$$k = \frac{G_Q - G_Z}{G_Q} \tag{8}$$

For liquid pipelines, the flow in the pipelines is expressed by the following equations.

$$\frac{\partial}{\partial t}(\rho A) + \frac{\partial}{\partial x}(\rho w A) = 0 \tag{9}$$

Where A is the cross sectional area of the pipeline.

$$\frac{\partial w}{\partial t} + w\frac{\partial w}{\partial x} + \frac{1}{p}\frac{\partial p}{\partial x} + g\sin\alpha + \frac{\lambda w^2}{2D} = 0 \tag{10}$$

Where g is the acceleration due to gravity, α is the inclined angle of the pipeline to the horizontal, x is the length of the pipeline from the inlet to the outlet, and w is the velocity. The fluid flow in a pipeline without leak is given as:

$$\frac{p_Q - p_Z - \rho g(h_Z - h_Q)}{w^2} = \frac{\lambda \rho L}{2D} \tag{11}$$

Where h_Z and h_Q are elevations of the inlet and the outlet ends of the pipeline.

The equation for a leaking pipe along the C_*^- path and C_*^+ path in Fig. 8 is represented by Eqs. (12) and (13) as:

$$\frac{p_Q - p_L - \rho g(h_L - h_Q)}{w_Q{}^2} = \frac{\lambda \rho L_r}{2D} \tag{12}$$

$$\frac{p_L - p_Z - \rho g(h_Z - h_L)}{w_Z{}^2} = \frac{\lambda \rho (L - L_r)}{2D} \tag{13}$$

Where w_Q and w_Z are the velocities of the upstream and downstream of the leakage point and h_L is the elevation of the leakage point.

The leak rate at the leak point is determined as:

$$k = \frac{w_Q - w_Z}{w_Q} \tag{14}$$

The wall of the pipe introduces loss of pressure in the fluid flow due to the nature of the pipeline internal surface. This loss which is known as friction factor, or the flow coefficient is determined using the Darcy Weisbach equation. The Darcy–Weisbach equation is an equation, which relates the pressure loss or the head loss due to friction along a given length of pipeline to the average velocity of the fluid flow for an incompressible fluid. The friction factor for fluid flow in pipelines in determined using the Darcy Weisbach equation or the Moody diagram and the Colebrook equation.

Given a cylindrical pipe of uniform diameter D, the pressure loss due to viscous effects Δp is proportional to length L and can be characterized by the Darcy-Weisbach equation shown in Eq. (15).

$$\frac{\Delta p}{L} = \frac{\rho}{2} f_D \frac{v^2}{D} \tag{15}$$

Where f_D is the Darcy friction factor (also known as the flow coefficient).

For laminar flow, the friction factor is replaced by the following expression:

$$f_D = \frac{64}{Re} \tag{16}$$

There are numerous models to calculate Darcy's friction coefficient, or friction factor. Most of them are empirical models limited to the range of experimentation in which they were formulated, such as Moody equation. However, Colebrook-White model has been widely preferred due to its precision within turbulent conditions.

The Moody Diagram shown in Fig. 9 is a log-log plot of the Colebrook correlation on axes of friction factor and Reynolds number, combined with the $f = 64/Re$ result from laminar flow.

7. Leak localization

Leak localization is the process of determining the leak point whenever the leak occurs. This is required to enable operators deploy appropriate strategies to contain the spills from the pipeline and minimize the environmental impact of the leak.

Leak localization methods include: gradient intersection method, wave propagation method and extended wave propagation method.

These methods can be deployed either independently or combined to achieve higher accuracy in the localization of pipeline leaks.

7.1. Gradient intersection method

The gradient intersection method uses the pressure profile along the pipeline to localize the leak. The pressure drop for a horizontal line without any elevation is linear as such if a leak occurs, the flow before the leak site increases and decreases after. This results in an increase in the pressure drop before the leak and decreases after the leak generating two lines with different gradients. Following these lines to the intersection, the leak site can be determined. While this method is easy to deploy, it has several limitations which limits its application. The system can be used to localize spontaneous and creeping leaks can be localized with a good accuracy in stationary operation. The accuracy of the system depends on the total length of the pipeline and this localizing accuracy is poor when the pipeline is in transient operation. The following parameters must also be factored in the method to achieve a usable result. These include changes in the height, cross-section, and pipe friction along the pipeline.

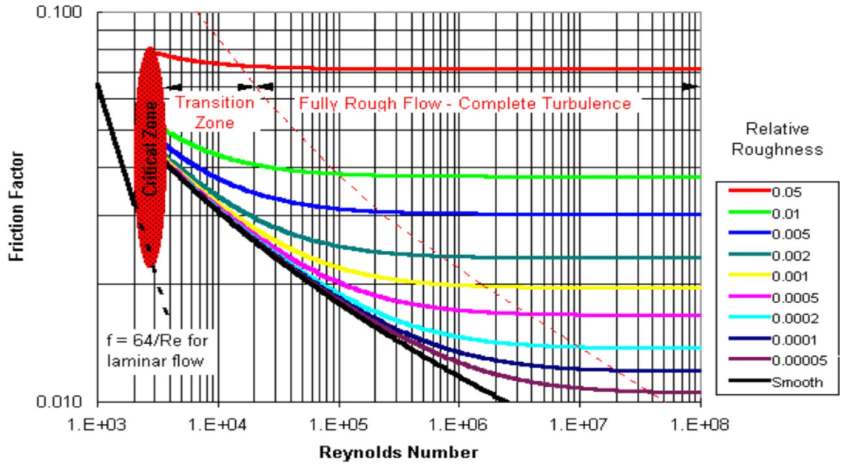

Fig. 9. Moody diagram (Moody, 1944).

7.2. Wave propagation method

This method uses the sound velocity of the leaks in the pipeline. Pipeline leaks create a negative pressure wave which propagates in both directions of the pipeline at the speed of sound. Pressure gauges at the inlet and outlet record these pressure waves and their time of arrival at the sensors. The time of arrival of the leaks at the two sensors can be analysed to determine the leak points. The arrival of the pressure waves at both sensor at the same time would indicate that the leak point is at the midpoint on the pipeline assuming a uniform density in the hydrocarbon and same speed in both directions. The wave propagation method achieved good accuracy both during stationary and transient operation provided that the operational pressure waves are compensated for. The method can be used during pumping and during pauses in pumping. One key limitation of this method is the fact that creeping leaks and spontaneous leaks that are not large enough cannot be detected with this method. The pressure gauges at both ends must be able to measure the arrival time of the pressure wave as accurately as possible.

7.3. Extended wave propagation method

This system increases the accuracy of the wave propagation method by using multiple sensors. When a leak occurs, the pressure waves can reach the gauges/sensors faster. By considering the sensor sampling time and the actual fluid density / sound velocity profile the exact point in time at which the pressure wave reached the sensors can be narrowed down even further thus improving the resolution of the leak point localization.

8. Components of RTTM

The RTTM which has the best leak detection performance (Lunger and Karami, 2019) relies on data from the field sensors which are fed into the model to both determine the commencement of a leak and the location of the leaks. The components of these system include the field-based sensors clamped on the pipeline, the data transmission system, and the Transient model. These are briefly explained in the next section. These are explained in the following subsections.

8.1. Clamp-on sensors with data transmission capacity

The use of clamp-on sensors is an approach where sensors are clamped on the pipeline at selected points to monitor parameters of the pipe at those locations (Shao et al., 2019). These sensors monitor the vibration and temperature on the pipeline. These measurements can be used to detect the flow in the pipe and the presence of leaks. The sensors are connected to microcontrollers and transmission system for the transmission of the acquired data to the cloud storage location. Specially developed algorithms are then used to analyse the data and determine the commencement of a leak, the localization, and the quantifying of the leak (Turner and Mudford, 1988). The proposed system diagram for the RTTM system and the sensors are shown in Fig. 10.

Fig. 10 shows the network architecture of the proposed RTTM based leak detection system with integrated field sensors. The field sensors are responsible for acquiring the vibration and temperature data from the section of the pipeline where they are installed. The sensors transmit the acquired data to the cloud location where machine learning algorithms are utilized to extract the required information for the leak location. The status of the lines can be monitored in real-time by all the responsible parties. Due to the large volume of data generated from the leak time detection sensors, dimensionality reduction algorithms must be engaged to reduce the volume of data utilized in the algorithms. Key challenges associated with large volume data transmission include shorter battery life and the presence of anomalies which extend the processing time for the data processing algorithms.

8.2. Impact of local geography

The local geography around the pipeline RoW has an impact on both the detection time for the leaks and the rate of spread of the oil spills into the environment. One key reason for this is the impact of the hydrostatic pressure on the fluid flow in the pipeline and the impact of the natural environment on the pipe. Leaks from pipelines in swampy locations are difficult to localize especially when the pipe is on the river floor. The effect of the water load also alters the pipeline pressure differential as the hydrostatic pressure from the water head reduces the leak pressure and this can cause the pressure due to the leak to fall within the pressure threshold of the pipeline there by making it difficult for the leak detection system to use pressure profile to determine the leak presence and leak location. Surface pipelines are easier to manage when it comes to pipeline leak detection and containment as the leaks are visible and there is no hydrostatic pressure to interfere with the pipeline pressure profile readings.

For pipelines installed in very cold locations such as North Dakota, the formation of ice can cause flow assurance and process safety issues, such as restricting flow path, pipeline plugging, failure of pipeline components, the release of hazardous liquid, and fire (Xu et al., 2018). The presence of snow cover on the pipeline and the installation of the sensors in buried pipelines increases the transmit power required for the transmissions from the sensors to get to the office domain IT infrastructure. Other topological environments include mountainous locations re-

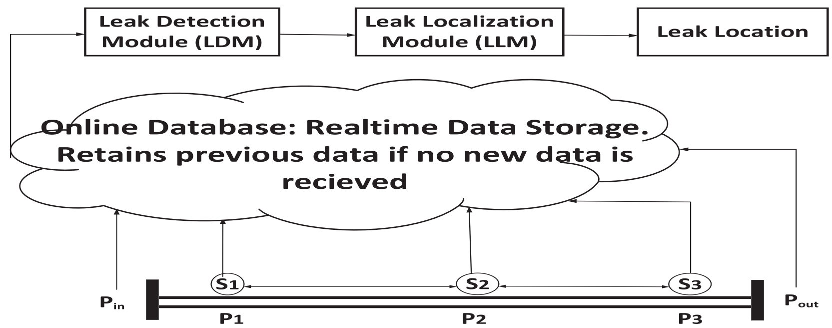

Fig. 10. Pipeline leak detection system using field-based sensors and RTTM.

quiring steep pipeline installations, areas with high vegetation density and locations that are difficult to access. All these locations introduce challenges for pipeline leak detection and localization. While some of the environmental parameters can be factored into the design, access to those location in the event of a pipeline leak would be both costly and time consuming.

8.3. Exception based transmission

The use of exception-based transmission will enable the transmission of only the leak related data. This approach creates a distributed processing system where each node can determine the pressure data to transmit. This is accomplished using a pressure threshold for each of the sensors and each sensor with a limited data processing capacity can monitor and compare the received data with the pre-set threshold and send the values that fall outside the pre-set window. This mode of transmission will reduce the volume of transmissions made from the sensors and limit the battery drain thus enabling the sensors to transmit at higher power to overcome the snow cover and the soil compaction over the pipeline. Fig. 11 shows the end-to-end configuration of the system from the sensors on the pipeline to the database and the leak detection and localization algorithms.

This model of real time data acquisition from sensors is suitable for installations in hard to access locations and for low-cost deployments. The current battery technology coupled with the exception-based transmission can enable the battery life to extend beyond a year. The nature of the sensor installation makes it suitable for surface pipeline installation however, it can be used for underground pipelines if the sensor sections are provided with suitable housing for easy access during battery replacement and providing a reduced attenuation for the signal transmission from the sensors located on the pipeline.

The methodology for this work includes the development of the pipeline model using the Darcey Weisbach equations, the Colebrook equation and the parameters of the pipeline and the fluid flowing in the pipeline to determine the pressure profile of the pipeline. These data will be compared with the pressure profile of a similar pipeline with the same parameters and both for the pipe, the topology and the fluid being transported. In view of the challenges associated with acquiring actual pipeline pressure profile, simulation software such as PIPESIM (He et al., 2017) will be used to generate the pipeline pressure profile with multiple leak points. The use of the simulation software provides an opportunity to model realistic pipeline installation environments, select the desired pipeline specification and simulate various leak scenarios (Lunger and Karami, 2019). MATLAB and EPANET are other simulation software that have also been with very promising results for the detection of pipeline leaks (Carbó-Bech et al., 2017; Candelieria et al., 2014). The detection of leaks from these pipelines will be achieved by comparing the pressure values at each point both in the model generated value and the simulation generated value or actual pipeline pressure values. A threshold value is added to the simulation value to cater for spurious

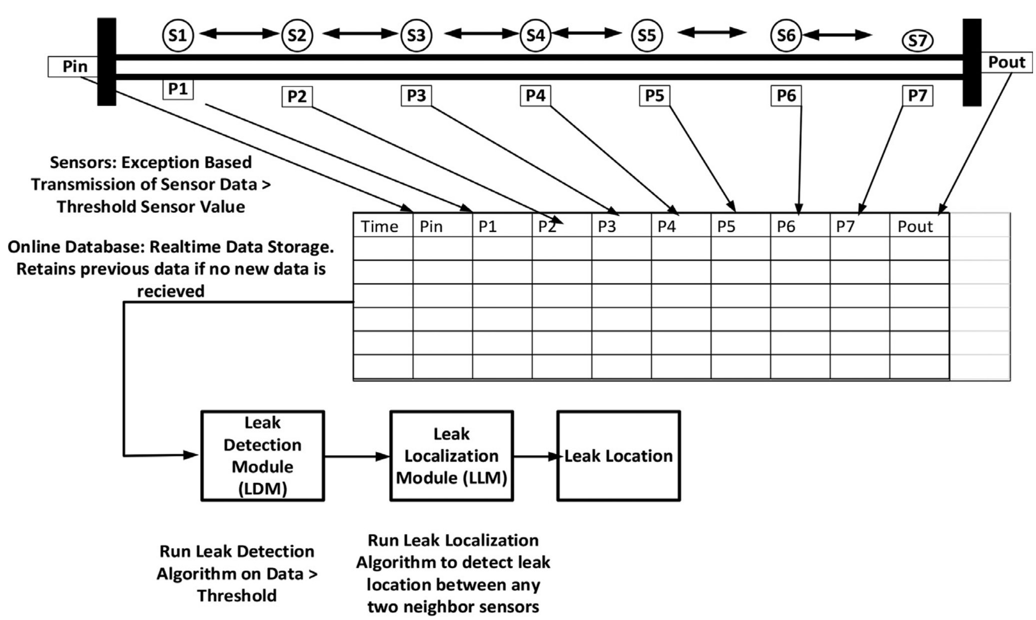

Fig. 11. End-to-end architecture of the proposed leak detection system.

pressure bursts occasioned by the irregularities of the pipeline internal surface. A leak is assumed to occur when the model value exceeds the sum of both the threshold value and the simulated leak value. The pipeline is broken into sections bounded by the pressure sensors and the leak detection process is applied to each section with the input pressure of each section computed as the current pressure expected at that point less the leak pressure experienced in the preceding pipe section.

9. Data analytics applications for leak detection

Data analytics tools have also been utilized in the development of the leak detection algorithms with different success levels. The Gradient Boosting, Decision Trees (DT), Random Forest (RF), Support Vector Machine (SVM) and Artificial Intelligence (AI) have been utilized by researchers with varying levels of sensitivity, accuracy, and reliability. The key finding of the research suggests that data analytics and artificial intelligence can be utilized with the RTTM to improve the leak detection results (Akinsete and Oshingbesan, 2019; Kang et al., 2018). The application of Machine learning (ML) for leak detection relies on the analysis of data collected by the real-time sensors installed on the pipelines without using any simulation (Romano et al., 2011).

Artificial Neural Network (ANN) based approaches have been proposed by Caputo and Pelagagge (2003) and, more recently, by Sivapragasam et al. (2007). Their systems use pressure and flow to infer the leak location and severity, through an ANN trained on a dataset generated either by a mathematical model of the network or the hydraulic simulation software EPANET (Candelieria et al., 2014). Mashford et al. (2012) proposed the combination of the EPANET-based leakage simulation and machine learning by using the SVM. The SVM model has been trained on a dataset of leaks simulated on the junctions of the water distribution networks (while most approaches simulate leaks on pipes). The trained SVM classifier can infer the leaky junction(s) according to the pressure and flow values (Candelieria et al., 2014).

Other supervised machine learning approaches have been recently proposed, such as Genetic Programming (Wang et al., 2012), Bayesian approaches (Poulakis et al., 2003a; Li et al., 2006) and Hidden Markov Mode-based agents (Nasir et al., 2010). Table 4 presents a summary of key findings from the use of different data mining strategies and models for leak detection.

10. Selected case studies

This section presents discussions on selected cases studies of Data analytics applied to leak detection

10.1. Olga simulations

This case study focuses on determination and analysis of leak estimation parameters in two-phase flow pipelines using OLGA multiphase software (Vandrangi et al., 2021). This focus of the paper was on the leak detection parameters which include mass flow rate, temperature, and pressure with and without leak in the pipelines. These parameters were evaluated using the OLGA multiphase software. An OLGA based computerized model was used in the leak simulation for analysing inlet and outlet parameters such as mass flow rate, temperature, and pressure over the flow inside the pipeline. The leak sizes were varied from 0% to 50% leak opening and the inlet and outlet parameters were measured and studied. The key findings of the research include the following: The pressure, mass flow rates are observed to decrease with increase in leak size, while temperature decreases with leak size until 25% and later increases. From the research, the Mass flow rate was observed to be the most important parameter in detecting a leak and localizing it. The maximum percentage of variation in mass flow rate was observed as 33.6% for 50% leak openings, for single leak and 32.4% for multi leak scenario. Another paper which focuses on the use of the Olga simulation is Lunger and Karami (2019).

10.2. Leakage location

This was an integrated study for detection and location modelling for leakages in liquid pipelines (Liu et al., 2016). This work focused on the development of an integrated model for leak detection and localization capable of detecting background leaks and very small microleaks in liquid pipelines. The model implemented a dynamic monitoring module (DMM) for detecting large leakages using amplitude propagation and attenuation model. It also utilizes a static testing module (STM) which is based on the pressure loss model to detect micro-leakages. The results show that the integrated model can detect nearly all leakages. For the DMM, the smallest detected ratio of leakage orifice to pipe diameter (RLOPD) in the field is 1/41.4, with location errors on the order of 1%. For STM, the smallest detected leakage rate is just 0.0044%/h in the field. Thus, the model has capacity to both detect and locate leakages from liquid carrying pipelines.

10.3. Liquid leakage

This study introduced a method for simulating the entire leaking process and calculating the liquid leakage volume of a damaged pressurized pipeline (He et al., 2017). For this work, the authors utilized three models to determine the leaking flowrate and volume. The negative pressure wave propagation attenuation model was used to calculate the sizes of the leak orifices. The transient oil leaking model which comprised of the continuity, momentum conservation, energy conservation and orifice flow equation was used to calculate the leakage volume. The third model was the steady- state leaking model and it was utilized to calculate the leakage after the valves and the pumps have been shut down. Validation of the numerical simulation was done using two types of leakage test with different sizes of leakage holes and Sinopec product pipelines. The leaking process under difference conditions were also described and analysed. The authors also utilized the Synergi Pipeline Simulator (SPS) software to simulate the pipeline and undertake the studies. The results obtained includes the following: The models can be applied to predict the equivalent diameter of the leaking orifice, the leakage volume during the unsteady leaking process and the ultimate volume of the steady leakage. The errors observed were within 7.6%. This study has a wide application area as it has the potential to provide guidance for estimating economic loss, evaluating the influence of accidental oil leaking, and designing remediation technology for containing the leaks. The models can also be widely applied to other liquid-transporting pipelines.

10.4. Leakage monitoring

This leak detection and location model (Liu et al., 2016) was based on the amplitude attenuation model of the dynamic pressure waves and was used for the detection of gas leaks. This was compared with the traditional method based on the propagation velocity and time differences as determined by the waveforms of the upstream and downstream signals. The results indicate that all leakages can be detected by both methods but that the largest location error of the traditional method is −0.780%, whereas the largest location errors with respect to the new method are 0.054%. It is further determined that the influence of the gas flow effects cannot be ignored by either method. The conclusions drawn suggest that the proposed methods can be applied to monitor gas pipelines. The experimental layout of the leak detection system is shown in Fig. 12.

10.5. Two-point leakages detection

Fu et al. (2021) proposed pressure distribution analysis for the detection of the leak points. This was done using experimental studies and

Table 4
A summary of Data Analytic models for leak detection.

Method	Authors	Approach	Key findings
Artificial Neural Network	Caputo and Pelagagge (2003)	The authors proposed a method of using ANN to estimate the leak location in piping networks. They generated input data for leaking and non-leaking states and used two neural networks in their proposal. The first NN was used for identifying the leaking branch and the second NN was for estimating the leakage amount and location	The results obtained showed that the leaking branch could be correctly identified with the leak size estimated to be between 2%– 10% of the actual value. The location of the leak could be estimated within 50–100 m of the actual leak location
	Mounce et al. (2010)	Simulated bursts on pipeline networks. Two sensor locations were used: one at the input of the network and one at the output. The sensors measured both pressure and flow. Five different burst locations were simulated	They were able to locate the bursts with an accuracy of 98.33%
	Salam et al. (2014)	Investigated an on-line monitoring system to detect leakages in pipe networks. They used pressure measurements at each junction as input data. The input data were generated by simulating leaks in the network	They used a Radial Basis Function Neural Network which could detect the leak location and sizes with an accuracy of 98%
	Zhang et al. (2016)		(1) The configuration of the neural network is critical for constructing an effective data driven model for a given system (2) ANNs rely too much on training samples. The ANN prediction accuracy is poor if the size of the training samples is small. To achieve adequate accuracy, sufficient training samples will be needed (3) The ANN training time increases as the size of the training samples increases (4) ANN needs to be retrained when the physical conditions of a water system change; this is quite common in the developing countries
	Romano et al. (2011)	Proposed a fully automated data-driven methodology using all the pressure and flow measurements available. This approach combined the use of an ANN for the short-term forecasting of hydraulic values and statistical processes to determine whether an abnormal event had occurred	The results obtained showed the potential of data-driven technologies for near real-time incident reporting
	Mounce and Machell (2007)	Used two ANN architectures (static ANN and time delay ANN) to detect the occurrence of bursts using flow data	The use of ANNs showed potential for identifying changes in the flow that corresponded to unusual fluctuations of this hydraulic variable
Support Vector Machines	De Silva et al., 2011	Investigated support vector machines used as pattern recognisers to detect leaks in pipe networks. They started with a SVM (Support Vector Machine) as a regressor to try and predict emitter coefficients. Six monitoring nodes were used to act as sensor locations. They selected 10 candidate leaking nodes and generated a data set with varying emitter coefficients	The SVM could, after training, achieve a testing accuracy of 76.8%. They then used 40 candidate leaking nodes and created a data set, for which a testing accuracy of 57.2% was achieved. They found that the predicted leak location was within 500 m of the actual leak location in all cases for a network that could fit into a 1,000 by 1,100 m square box. They went on to investigate whether the SVM could detect small leaks in the network. The smallest leak registered by EPANET to generate a pressure difference was a leakage of 90 l/hour. A new data set was created to which the SVM was trained. A testing accuracy of 35% was found
Bayesian Probabilistic Framework	Poulakis et al. (2003a)	Investigated the ability of a Bayesian probabilistic framework to detect leaks in a water pipe network. The derivation starts by assigning θ as the parameter to be optimized. This parameter includes the leaking pipe, location, and size of the leak. A network consisting of 50 pipes, 31 nodes, and 20 loops was used. The network forms a grid network supplied by one reservoir with one leak. They went on by introducing variation in pipe roughness coefficients, variation in the assumed demands, and a variation in the model measurements	They found that when the model measurements had an uncertainty of 2%, the location could be calculated. If the uncertainty in the model measurements was increased to 5% the model was unsure about the actual leak location
	Zhou et al. (2011)	Proposed a feasible Bayesian reasoning approach based on a recursive algorithm for leakage detection. The recursive algorithm is used to update a belief rule-based expert system, which can learn the relationship between the flows/pressures in monitoring points and leak sizes	The key findings from the Bayesian methods include the following. The characteristics of the Bayes method uses probability distribution to represent all forms of uncertainty, and the process of learning or reasoning is implemented through probability. This can avoid the unavoidable uncertainties in the process of measuring and modelling. However, the method has some disadvantages. (1) Assumptions need to be made for the probability distribution of training data. The recognition accuracy is usually very low when the assumed probability distribution is different from the real probability distribution (2) It is difficult to accurately estimate the variance matrix of the Bayes algorithm (3) Estimating probability density through training samples is a highly ill posed problem when the size of sample is small. Thus, to achieve better recognition accuracy the Bayes method usually needs a large size of training samples

(*continued on next page*)

Table 4 (*continued*)

Method	Authors	Approach	Key findings
	Costanzo et al. (2014)	Developed a new Bayesian calibration methodology for leakage detection and calibration. The network is first divided into several zones according to piping roughness, and then the roughness is calibrated using the Bayes method. The leakage zone is identified by comparing the roughness of a normal condition and the roughness of an abnormal condition	
	Romano et al. (2010)	Applied wavelet analysis, ANNs, statistical process control, and Bayesian inference systems and integrated them into a unified framework for burst/leak detection. The method was tested in a real-life district metre area of the United Kingdom, and the results illustrated that the method could respond quickly to bursts/leaks	
Random forest	Huang et al. (2018)	Proposed the use of a random forest classifier to detect bursts in real time by analysing successive time windows (every 15 min) of flow data from the pipeline network	These contributions demonstrate the use of a supervised learning technique for the detection of bursts in pipelines; however, all the applications to date have been focused at a DMA level and using SCADA flow (or pressure) data, which are often available in intervals between 1 min and 15 min
	Bohorquez et al. (2020)	Presented a technique that uses ANNs to predict the presence of different features (leaks and junctions) in a pipeline after the generation of a controlled transient event	The results demonstrated the potential of combining transient-based techniques, with ANNs to interpret the pressure traces

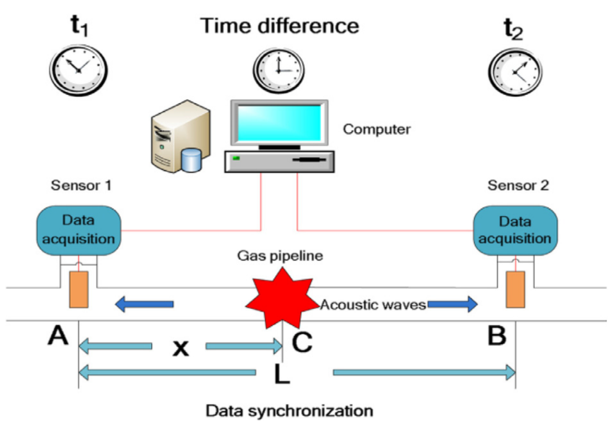

Fig. 12. Gas leak detection layout. The inlet and outlet are A and B respectively. *L* is the length of the pipe section and C is the leak point (Liu et al., 2016).

Computational Fluid Dynamics (CFD) simulation. Different dimensionless variables, which include the dimensionless leak location, leak rate and pressure drop were applied to the analysis. The locations of the leaks were detected using mathematical modelling. Multiple flowrate testing was conducted to detect the location of two leaks. The outcome of the research was that the method can be utilized to detect leakages under different flow rates. Fig. 13 shows the experimental layout of the research. Key findings from this research include the fact that different leak locations can lead to different pressure drop through the leaking pipeline. It was also observed that the pressure drop between the inlet and outlet of the leaking pipe is positively related to the inflow rate. When the leak locations are the same, the pressure drop through the leaking pipe becomes more significant with a larger inflow rate. The research also validated the use of simulations in leak detection studies as the parameters obtained in the simulation using the CFD software were of the same values as the those obtained from the experimental set up.

10.6. Single leakage diagnosis

In this study (Fu et al., 2020), a pipeline leak detection method was developed using the pressure distribution through the leak. A flow loop was used to undertake pipeline leak detection tests. Through the experiments, four parameters which are pressures and flow rates of pipeline inlet and outlet are recorded. The dimensionless analysis of these parameters was used to detect the leak location. To generate the solution to locate the leak, three dimensionless variables which are the dimensionless leak location and leak rate, and the dimensionless pressure drop were used. A 3D computational fluid dynamics (CFD) simulation was undertaken with a commercial software (FLUENT) to determine the leak points. The pressure distribution in the pipeline with leakage were verified using experimental data. Mathematical models were also developed to detect and evaluate the leak through the pipelines. The key findings from the CFD simulation results show that both the leak rate and location have significant effects on the pressure distribution through the pipe. This finding is identical to the outcome from the experiments. The mathematical model which is based on these dimensionless variables can be applied in locating the leak point in the real accident. The general conclusion from this research shows that simulation and experiments can be used to both determine the leaks and locate such leaks with a high degree of accuracy.

10.7. Real-time sensor data analysis

Oliveira et al., (2018) in this study, utilized data Science and advanced data analytics technologies to develop a leak detection system in a slurry pipeline. The techniques used to detect leakage were based on artificial intelligence, a machine learning model for the energy balance of the pipe combined with an anomaly detection technique approach. The pipeline was divided into sections and the system predicts energy at one point of the pipe based on another point. This is done to determine if there is a leak in that section. The machine learning model used in this work is the simple parametric linear regression model. The line used in this study had a length of 5.5 km and had two pressure gauges (PITR and PIT1) installed after each pump assembly as shown in the Fig. 14.

Other gauges installed on the line include a flow metre (FIT) and other density (DIT). Along the tailings line, there are two pressure gauges: PIT2, which is 3.3 km of pump sets and PIT3 which is 4.5 km from the beginning of the pipeline. The values measured by these sensors are shown in Table 5.

The pipeline system and the working fluid was modelled based on the energy conservation principle and the linear regression algorithm

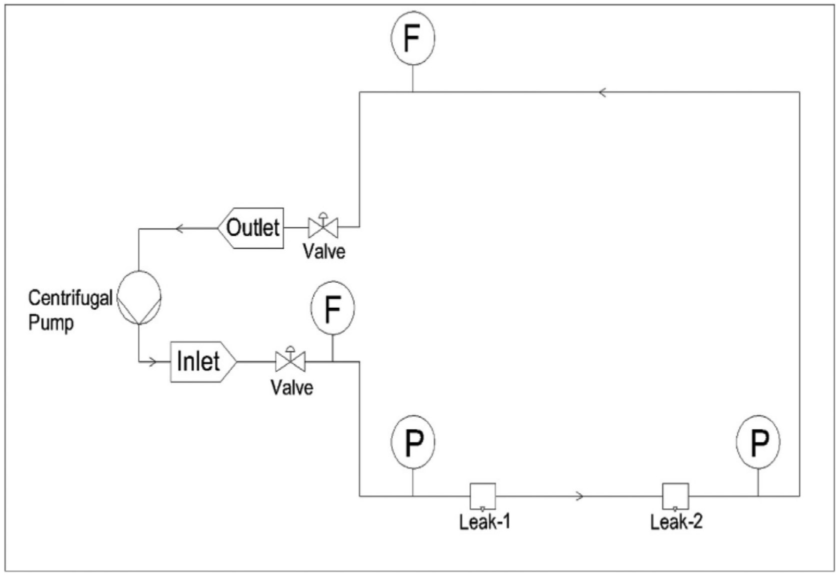

Fig. 13. Experimental set up (Fu et al., 2021).

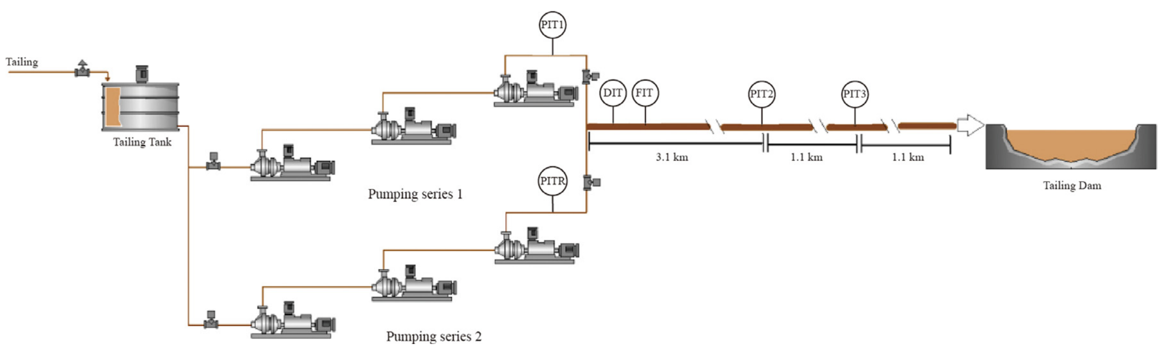

Fig. 14. Pipeline overview. (Oliveira et al., 2018).

was used to predict the energy at different points on the pipeline. The energy measured along the pipeline under normal operations without leaks was used to train the machine learning model. This enabled the model to determine the normal behaviour of the line or the expected energy balance of the line. One key challenge with this approach is the lack of leak data for training the model, thus it would be difficult to train the model for leak conditions.

The anomaly detection technique is the most suitable approach for training the model as it helps to define the threshold between anomalous and not anomalous behaviour. The leak in this case is an anomalous operation and is defined as the difference between the mean and the standard deviation of the error the reading of both the leak and the no leak models. The system was designed to activate protection systems that shut off the lines whenever a leak was detected, this meant that false positives had to be kept to the barest minimum as it would lead to production losses. Fig. 15 shows the response of the leak detection model. The real-time sensor data values and the model estimates align when there is no leak but when an anomalous behaviour is detected, the leak alarm is triggered.

Table 5
Reading from the sensors installed on the pipeline (Oliveira et al., 2018).

Sensor	Mean Value
PIT1	1,441.6 kPa
PITR	1,412.2 kPa
PIT2	755.1 kPa
PIT3	931.6 kPa
DIT	1.5 g/cm^3
FIT	1,000 m^3/h

One key challenge with this solution is the definition of the threshold between normal or abnormal behaviour of the pipeline line. This challenge exists the position of the threshold impacts on both the precision and robustness of the system and determine the detection of leak volume and detection time. The lower the threshold, the greater the accuracy, but this also means the greater the number of false positives which leads to pipeline shutdown and production losses. This challenge can be overcome by tuning the model to the operator facilities to factor in the impact of the operational sequences and terrain in the development of the model.

The implementation of the system used a very simple architecture. The data, which was continuously uploaded onto the plant historian, was read from there and stored an updated in a SQL database every five minutes. The analytics platform (KNIME) reads such data (sensor signals) and interprets it using data mining techniques (cleaning and processing of data), and machine learning techniques (linear regression and anomaly detection). The results (which are leak alarms) are written in the same SQL database. The alarm signals are also made available via OPC to the plant supervisory system.

Three leaks' tests were performed on a real pipeline to validate the leak detection system. The system was able to detect the leaks. When the pre-set thresholds were exceeded, the leak alarms were triggered as shown in Fig. 16. The blue line represents the real time measured data

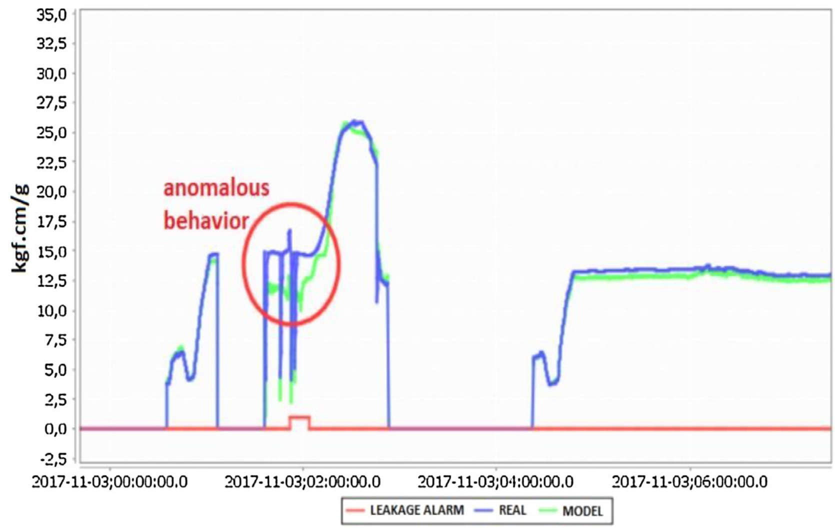

Fig. 15. Real time leak detection system (Oliveira et al., 2018).

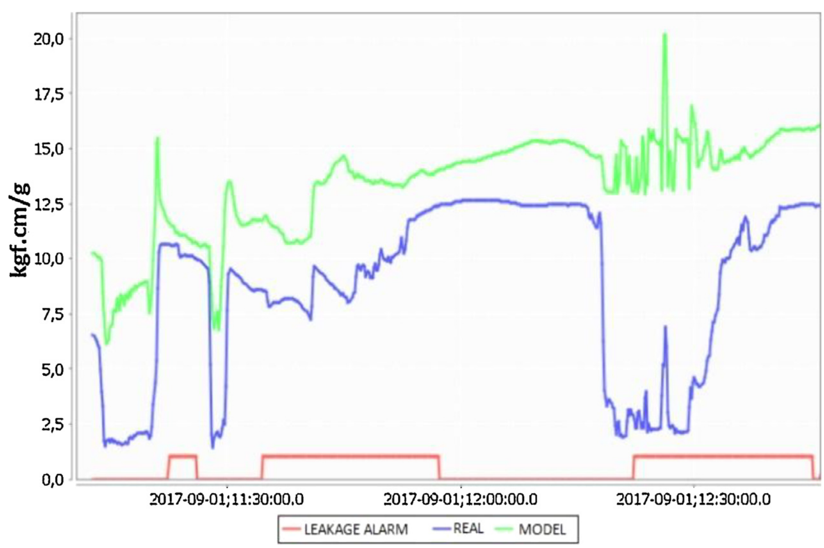

Fig. 16. System validation tests (Oliveira et al., 2018).

while the green line represents the model estimates for that section of the line. When there is an unusual change or reduction in the in the real-time measurements as compared with the model values, the leak alarms are triggered.

One key challenge posed by the system during the tests was the number of false positives. While the system was able to detect small leaks, it had a high number of false positives per day (5 false positives per day). The solution to this was a trade-off between robustness and accuracy the threshold of the system was re-tuned using the leak data collected during the tests. This resulted in the retraining of the model and the use of a higher threshold for defining leak scenarios. This retraining and remodelling of the system resulted in the detection of leaks but with fewer false positive detections (5 false positives per month). The system proved to be able to detect leaks using the real time data acquired from the pipeline and the dynamic model of the line.

11. Conclusion

Leak detection remains as a very critical area in oil and gas operations as such the need to develop robust leak detection and localization systems continue to be of prime importance to the oil and gas operators. Data analytics provides an opportunity for the application of intelligent algorithms for the analysis of the data acquired from the sensors located in the field for the detection and localization of pipeline oil and gas leaks. Mathematical models and simulation software have been shown to provide comparable results with experimental data with very high levels of accuracy. These systems also provide the opportunity for the development of early warning systems for pipeline leaks by analysing the vibration signature of the pipeline and identifying variations caused by corrosion of the pipelines before eventual failure and oil spill occurs. The vibration signature can also be studied to determine the state of the pipeline internal condition and detect the presence of hydrate formation on the pipeline internal surface. Several machine learning models, and algorithms have also been used for detecting and locating leaks pipeline pressure profile data. The most prominent of these models are the Artificial Neural Network (ANN) and the Support Vector Machine (SVM). Neural networks and data from leaking and non-leaking pipelines were utilized to determine the leakage amount and the location of the leak. In the absence of pipeline data from operators, simulation data was used for the studies with pressure being the most predominant pipeline parameter used for leak detection. The key findings from the use of Artificial Neural networks includes the fact that ANNs rely too much on training samples, its prediction accuracy is poor if the size of the training samples is small. To achieve adequate accuracy, sufficient training samples will be needed however, the ANN training time increases as the size of the training samples increase. SVM was used as a pattern recognition system to detect the variations in the sensor data generated from the pipelines and it was able to detect small leaks. The use of machine learning has been shown to play a critical role in the detection of leaks. The ability to develop leak detection models using synthetic data (data generated from pipeline models) increases the suitability of the use of machine learning for leak detection and localization algorithms. While some of the models require large volumes of training data, threshold detection can be combined with the machine learning algorithms to enable the development of a robust machine learning enabled leak detection system.

Authors' contributions

Francis Idachaba: Undertook the research and wrote the paper, Minou Rabiei: Made substantial contribution to the conception of the work, provided review, and approved the publication.

Declaration of Competing Interest

No conflict of interest is associated with the work.

References

Carbó-Bech, A. , De Las Heras, S.A., Guardo, A., 2017. Pipeline leak detection by transient-based method using MATLAB R functions. doi:10.20944/preprints201706.0007.v1

Candelieria, A., Contib, D., Archettia, F., 2014. A graph-based analysis of leak localization in urban water networks. In: 12th International Conference on Computing and Control for the Water Industry, p. CCWI2013.

Akerkar, R., 2019. Machine learning. Artificial Intelligence for Business. Springer Briefs in Business. Springer, Cham doi:10.1007/978-3-319-97436-1_2.

Salam, A.E.U., Tola, M., Selintung, M., Maricar, F., 2014. On-line monitoring system of water leakage detection in pipe networks with artificial intelligence. ARPN J. Eng. Appl. Sci. 9 (10), 1817–1822 October 2014.

Begovich, O., Valdovinos-Villalobos, G., 2010. DSP application of a water-leak detection and isolation algorithm. In: 7th International Conference on Electrical Engineering Computing Science and Automatic Control (CCE), pp. 93–98.

Baroudi, U., Al-Roubaiey, A., Devendiran, A., 2019. Pipeline leak detection systems and data fusion: A survey. IEEE Access.

Begovich, O., Pizano-Moreno, A., Besancon, G., 2012. Online implementation of a leak isolation algorithm in a plastic pipeline prototype. Lat. Am. Appl. Res. 42, 131–140.

Bohorquez, J., Alexander, B., Simpson, A.R., Lambert, M.F., 2020. Leak detection and topology identification in pipelines using both fluid transients and artificial neural networks. J. Water Resour. Plann. Manage. 146 (6), 04020040 https://doi.org/10.1061/(ASCE) WR.1943-5452.0001187.

Caputo, A.C., Pelagagge, P.M., 2003. Using Neural Networks to monitoring piping systems. Process Saf. Prog. 22 (2), 119–127.

Cui, C., Jiang, S., He, X., Wang, K., Shao, H., Wu, Z., 2018. Experimental study on the location of gas drainage pipeline leak using cellular automata. J. Loss Prev. Process Ind. 56, pp 68–77.

Costanzo, F., Morosini, A.F., Veltri, P., Savi´c, D, 2014. Model calibration as a tool for leakage identification in WDS: a real case study. Procedia Eng. 89, 672–678.

Li, X., Wang, X., Zhao, X., Li, G., 2006. Bayesian theorem based on-line leakage detection and localization of municipal water supply network. Water and Wastewater Engineering 12.

Liu, C., Li, Y., Fang, L., Han, J., Xu, M., 2016. Leakage monitoring research and design for natural gas pipelines based on dynamic pressure waves. J. Process Control 50, 66–76 2017.

De Silva, D., Mashford, J., Burn, S., 2011. Computer aided leak location and sizing in pipe networks. Urban Water Security Research Alliance Technical Report 17,April 2011.

Oliveira, E., Fonseca, M. D., Kappes, D., Medeiros, A., 2018. Leak detection system using machine learning techniques. 6th International Congress on Automation in Mining.

Idachaba, F.E., 2016. Monitoring of oil and gas pipelines by use of VTOL-type unmanned aerial vehicles. Oil and Gas Facilities J. 2016.

Romero-Tapia, G., Fuente, M.J., 2018. Leak localization in water distribution networks using fisher discriminant analysis. Vicenç Puig 51 (24).

GlobalData, 2019. Global oil and gas pipelines industry outlook to 2023 – Capacity and capital expenditure outlook with details. of all operating and planned pipelines.

He, G., Liang, Y., Li, Y., Wu, M., Sun, L., Xie, C., et al., 2017. A method for simulating the entire leaking process and calculating the liquid leakage volume of a damaged pressurized pipeline. J. Hazard. Mater..

Fu, H., Wang, S., Ling, K., 2021. Detection of two-point leakages in a pipeline based on lab investigation and numerical simulation. J. Pet. Sci. Eng. 204, 108747 2021.

Fu, H., Yang, L., Liang, H., Wang, S., Ling, K., 2020. Diagnosis of the single leakage in the fluid pipeline through experimental study and CFD simulation. J. Pet. Sci. Eng. 193, 107437 2020.

Xu, H., Bbosa, B., Pereyra, E., Volk, M., Sam Mannan, M., 2018. Oil transportation in pipelines with the existence of ice. J. Loss Prev. Process Ind. 56, 137–146.

Huang, P., Zhu, N., Hou, D., Chen, J., Xiao, Y., Yu, J., et al., 2018. Real-time burst detection in district metering areas in water distribution system based on patterns of water demand with supervised learning. Water (Basel) 10 (12), 1765 https://doi.org/10.3390 /w10121765..

Kang, J., Park, Y.-J., Lee, J., Wang, S.-H., Eom, D.-S., 2018. Novel leakage detection by Ensemble CNN-SVM and graph-based localization in water distribution systems. IEEE Trans. Indust. Electron. 65 (5).

Fiedler, J., 2016. An overview of pipeline leak detection technologies. KHRONE Inc https://asgmt.com/wp-content/uploads/2016/02/004.pdf.

Moody, L.F., 1944. Friction factors for pipeline flow. In: Semi-Annual Meeting of the American Society of Mechanical Engineers, p. 1944 June 19-22.

Adegboye, Mutiu Adesina, Fung, Wai-Keung, Karnik, Aditya, 2019. Recent Advances in Pipeline Monitoring and Oil Leakage Detection Technologies: principles and Approaches. Sensors MDPI.

Mashford, J., De Silva, D., Burn, S., Marney, D., 2012. Leak detection in simulated water pipe networks using SVM. Appl. Artif. Intellig. Int. J. 26 (5), 429–444.

Mounce, S.R., Mounce, R.B., Boxall, J.B., 2010. Novelty detection for time series data analysis in water distribution systems using support vector machines. J. Hydroin. 13 (4), 672–686 https://doi.org/10.2166 /hydro.2010.144.

Nasir, A., Soong, B.H., Ramachandran, S., 2010. Framework of WSN based human centric cyber physical in-pipe water monitoring system. In: 11th International Conference on Control, Automation, Robotics and Vision, pp. 1257–1261.

Matthew L., 2013. Average 250 pipeline accidents each year, billions spent on property damage. https://www.desmog.co.uk/2013/04/05/average-250-pipeline-accidents-each-year-billions-spent-property-damage

Akinsete O., Oshingbesan A., 2019. Leak detection in natural gas pipelines using intelligent models. Nigeria Annual International Conference and Exhibition.

Physiker, R.T., 2003. Model-based pipeline leak detection and location. 3R International 42, 455–460.

Poulakis, Z., Valougeorgis, D., Papadimitriou, C., 2003a. Leakage detection in water pipe networks using a Bayesian probabilistic framework. Probab. Eng. Mech. 18, 315–327.

Sukarno, P., Sidarto, K.A., Trisnobudi, A., Setyoadi, D.I., Rohani, N., Darmadi, D., 2007. Leak detection modelling and simulation for oil pipeline with artificial intelligence method. ITB J. Eng. Sci. 39 (1), 1–19 B, No.2007.

Zhang, Q., Wu, Z., Zhao, M., Qi, J., Huang, Y., Zhao, H., 2016. Leakage zone identification in large-scale water distribution systems using multiclass support vector machines. J. Water Resour. Plann. Manage., 2016 142 (11), 04016042 ACSE.

Romano, M., Kapelan, Z., Savić, D., 2011. Real-time leak detection in water distribution systems. In: Water Distribution Systems Analysis, pp. 1074–1082.

Romano, M., Kapelan, Z., Savi´c, D.A., 2010. Real-time leak detection in water distribution systems. In: Proc., 12th Water Distribution Systems Analysis Conf. ASCE, Reston, VA.

Silva, R.A., Buiatti, C.M., Cruz, S.L., Pereira, J.A.F.R., 1996. Pressure wave behaviour and leak detection in pipelines. Comp. Chem. Eng. 20.

Seiders, E.J., 1979. Hydraulic gradient eyed in leak location. Oil Gas J 77, 739–744.

Sivapragasam, C., Maheswaran, R., Venkatesh, V., 2007. ANN-based model for aiding leak detection in water distribution networks. Asian J. Water Environ. Pollut. 5 (3), 111–114.

Mounce, S.R., Machell, J., 2007. Burst detection using hydraulic data from water distribution systems with artificial neural networks. Urban Water J. 3 (1), 21–31 February 2007.

Sun, L., 2012. Mathematical modelling of the flow in a pipeline with a leak. Math. Comput. Simul. 82, 2253–2267 2012.

Lunger, T., Karami, H., 2019. Leak detection in wet natural gas transportation within hilly terrain pipelines. In: Unconventional Resources Technology Conference, Denver, Colorado, USA, pp. 22–24 July 2019.

Vandrangi, S.K., Lemma, T.A., Mujtaba, S.M., Pedapati, S.R., 2021. Determination and analysis of leak estimation parameters in two-phase flow pipelines using OLGA multiphase software. Sustain. Comput. Informat. Syst. 31, 100564.

Turner, W.J., Mudford, N.R., 1988. Leak detection, timing, location and sizing in gas pipelines. CSIRO Division of Mineral and Process Engineering. Lucas Heights, Menai, NSW 2234, Australia.

Wang, L., Zhangq, H., Jia, H., 2012. A leak detection method based on EPANET and Genetic algorithm in water distribution systems. Software Engineering and Knowledge Engineering: Theory and Practice – Advances in Intelligent and Soft Computing 14, 459–465.

Shao Y., Li X., Zhang T., Chu S., Liu X., 2019. Time-series-based leakage detection using multiple pressure sensors in water distribution systems sensors MDPI.

Zhou, Z., Hu, C., Xu, D., Yang, J., Zhou, D., 2011. Bayesian reasoning approach based recursive algorithm for online updating belief rule based expert system of pipeline leak detection. Expert Syst. Appl. 38 (4), 3937–3943.