Measurement 166 (2020) 108262

Contents lists available at ScienceDirect

Measurement

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

Assessment of the structural conditions in steel pipeline under various operational conditions – A machine learning approach

Michel Saade, Samir Mustapha *

Laboratory of Smart Structures and Structural Integrity (SSSI), Mechanical Engineering, American University of Beirut, Beirut, Lebanon

ARTICLE INFO

Article history:
Received 1 March 2020
Received in revised form 14 July 2020
Accepted 15 July 2020
Available online 19 July 2020

Keywords:
Pipe condition monitoring
Structural health monitoring
Fiber optic sensing
Machine learning

ABSTRACT

Oil and water transport pipeline systems are susceptible to damage due to harsh environmental conditions and operational factors. Hence, ongoing maintenance and inspection are required. The development of continuous and reliable monitoring will ensure the safe usage of these structures and assist in the extension of their life spans. In this study, the monitoring and assessment of pipelines are performed using a network of Fiber Bragg Grating (FBG) sensors mounted in the longitudinal and circumferential directions pipelines. The sensitivity of these measurements to assess pipe pressure and flow variations, and leakage detection and localization were evaluated. Water, at a controlled pressure and flow rate, was pumped into the designed 6-m pipe testbed. Leakage was simulated by opening one of the four designated valves installed on the pipe. Support Vector Machine (SVM) algorithms were implemented, using the collected data, to assist in the prediction of the structural condition of the pipe under various operational conditions. Pressure variations inside the pipe highly impacted the amplitude of the measured strain, increasing it significantly up to 20%. A flow rate increase of 5 GPM had the inverse effect, resulting in a 5% decrease in the amplitude of the measured strain. A change of leakage hole size greatly influenced the measured signal, resulting in a 55% change in amplitude between a 2-cm^2 and a 5-cm^2 hole. To determine the leakage location, only the temporal aspects of the signal were affected, resulting in slight shifts in sensor response times. The developed SVM classifiers reached accuracies of 88% for flow rate classification, greater than 95% for pressure classification, and 100% for leakage size classification. The accuracy of leakage localization did not exceed 72%. These results are promising for the monitoring of the structural conditions related to leakage detection and localization, based on the patterns observed.

1. Introduction

Pipeline systems are remarkable for their practicality, efficiency, and cost-effectiveness in transporting large quantities of dangerous substances [1], such as crude oil and petroleum products [2], compared with other modes of transport. These pipeline systems can be used for transport over short distances (within refineries or between neighboring installations), as well as over long distances. Oil and gas products from hundreds of thousands of wells, many of which are located in remote and hostile areas, are transported through pipelines [3].

The deployment locations and large size of pipeline networks make this transportation method susceptible to various types of failures. Such failures may not only cause product loss but also environmental damage as well as hazards and potential threats to life [4,5]. Pipeline failures, depending on their location and the nature of the transported substance, can have many negative effects on the economy, the environment, and public health. A recent pipeline failure, for instance, spilled 1.4 million liters of bitumen mixed with a diluent in Keystone, North Dakota [6]. The diluent, which quickly evaporated, rendered the air toxic, and the bitumen remaining in the ground produced a foul odor. Similar incidents have also occurred, and their consequences have been intrusive and destructive. Some examples are the 120,000-liter spillage in Garfield County, Oklahoma, the 300,000-liter spillage in Miller Grove, Texas, and the 40,000-liter spillage in Alberta, Canada, which all happened in 2019 [7–9]. No spillage-related injury or wildlife impact were reported from these events [10,6].

Given the immense scale of the pipeline networks and the severity of possible leakage repercussions, the structural health monitoring (SHM) of pipelines is a challenging yet necessary task. The methods developed over time are strictly nondestructive, as they must be applied to installed and running pipelines. Several methods have been developed for this purpose [11].

* Corresponding author.
E-mail address: sm154@aub.edu.lb (S. Mustapha).

https://doi.org/10.1016/j.measurement.2020.108262
0263-2241/

In this paper, the operational and structural condition assessment of steel pipes using a sensor network of fiber optic Fiber Bragg Grating (FBG) sensors was performed. Fiber optic sensors (FOS) were chosen for this application due to their superb capability to provide measurements over long distances. Further, they do not require electrical installations, making the application of this method on previously laid pipelines plausible [12]. The main aims of this study were to (i) investigate the efficacy of FBG strain sensors for the monitoring of the structural health and operational conditions of pipelines, (ii) investigate the effect of the number and orientation of sensors needed for high prediction accuracy, (iii) develop a good understanding of the effects of flow and leakage conditions on the surface strain of pipelines, and (iv) develop learning algorithms, based on FBG data, for the automation of the monitoring process.

2. Literature review

The methods of inspection that rely on externally mounted sensors to detect pipe leakages are called sensor-based methods. They include negative pressure wave (NPW), ultrasonic, eddy current, and electromagnetic flux methods [13]. Their main advantages are quantification and application without the interruption of normal pipeline operation. Jin et al. [14] worked on localizing leakages in gas pipelines using acoustic and pressure sensors mounted on the surface of the pipe. The authors managed to detect a leakage in a 170-km pipeline with an error of 720 m. Zhu et al. [15] mounted piezoelectric (PZT) wafers on a pipe and localized leakages based on the NPW theory [16]. They managed to locate leakages with an error of 2% on a 55-meter testbed.

Methods based on intelligent pigs have also been developed for the inspection of pipeline systems [17]. These methods are particularly useful for detecting and localizing changes in pipe wall thicknesses. However, pipeline operations may need to be disturbed to perform these tests, and the monitoring process is not continuous. Mishra et al. [18] discussed the application of intelligent pigs using the magnetic flux leakage technique to assess the structural health of pipelines. Alnaimi et al. [19] worked on designing a PIG with enhanced performance over traditional pigging devices. These authors developed a prototype able to monitor pipes of different diameters, perform cleaning tasks, and reduce the effect of water hammers, which are sudden peaks in pressure that can harm the pipe.

Further, single measurement analysis methods which work by detecting peaks or dips in the measurement of certain flow conditions are widely used. Any sudden change in the flow rate or pressure inside the pipe can indicate a potential defect. These methods are easy to implement but lack quantification and localization capabilities. Akib et al. [20] showed how such methods work. Given that these methods rely on comparing current data to previous data collected during normal running operations, the set-up requires the gathering of data, which can be time-consuming. Therefore, they showed how data for these methods can be generated using simulations.

Other methods are based on the uncompensated volume balance method in which the sum of flows entering the pipeline and the sum of flows exiting the pipeline are compared. Depending on the size of the network, this difference can fluctuate within certain limits. If, however, the difference is noticeably large, it can indicate a leak in the network. Martins et al. [21] compared mass balance methods and NPW methods. They concluded that mass balance methods are more precise for the calculation of leakage rate, but acoustic methods can detect the leakage faster and localize it with better accuracy.

Model-based leakage detection methods [22,23] in which a model of the pipeline under study is simulated can accommodate the closing of valves or changes in other control parameters in the pipeline. Flow properties at certain locations in the network can then be compared to flow properties at the same locations of the simulated network. This method needs highly accurate inputs and a computer capable of delivering good results in a reasonable amount of time. The SHM of pipelines can be compared to that of bridges and other large-sized civil engineering structures. In such cases, the use of different types of sensor data or machine learning (ML) algorithms can be used to facilitate the SHM process [24,25].

Of the above-listed methods, the most effective leakage localization methods in a pipe during normal operation are the ones with externally mounted sensors. Two of these methods are particularly effective: the NPW method and temperature-based methods.

The NPW method is concerned with the NPW that is generated when a pipe starts leaking fluid (change in pressure). This NPW travels through the fluid in the pipe to neighboring pressure sensors that pick up this change in operational conditions [16]. The leakage is then localized using the time difference for the NPW to reach the sensors upstream and downstream of the leak location.

On the other hand, temperature-based methods are based on monitoring the temperature of the medium surrounding the pipe. Even though these methods generally yield better outcomes and higher precision than the NPW method, they are subject to noise and errors caused by weather abnormalities or third-party intrusions [26].

Some other methods were developed for leakage detection. One of these was developed by Spirin et al. [27] who developed a casing that is sensitive to petroleum hydrocarbons. It consists of a polymer that reversibly swells when in contact with certain hydrocarbons, such as gasoline. An FBG sensor is then coupled to this casing for the detection of strain variations. An obvious disadvantage of this method for long pipelines is that the leakage must be in the vicinity of the sensor for the leaked fluid to contact the polymeric coating.

Another leakage detection method was developed by Lee et al. [28] who coupled PZT wafers to fiber optic FBG sensors. This method can detect liquid leakages by sending an acoustic wave in the longitudinal direction through a fiber optic cable. The stretch of the fiber between the transmitter and the FBG then acts as a sensor. When a leakage occurs, the fluid that comes into contact with the fiber attenuates the ultrasonic wave traveling along that fiber. The FBG sensor then detects a decrease in amplitude, signaling a leak. Although this method is highly effective, it compromises the benefits of using an FOS, as it requires electrical wiring along the pipe to actuate the PZT wafer.

All the previously mentioned methods generate different types of data that must be analyzed to determine the structural health of the tested structure. Analysis can be done by an operator, but its automation can be achieved by introducing artificial intelligence and training ML models to identify and quantify various types of flaws in the system [29–31]. The automation of the detection process starts by finding a trait in the signal that points to a leakage/flaw. Khodairy et al. [32] worked on developing a classifier that detects metal loss and cracks with an accuracy of 97.7% by looking at the magnetic flux leakage data collected for the tested specimen. The authors also developed a classifier that can determine the severity of the material loss using the same set of data, reaching an accuracy of 98.3%. The application of ML has been extended to the detection of cracks in self-healing concrete structures [33]. An interesting application of ML is a deep neural network, which evaluates the flexoelectric effect in truncated pyramid nanostructures under compression [34]. While applying ML algorithms, it

is very important to obtain well-balanced data for robust classification. Given that large amounts of abnormal data may not be available (specifically in real structures), the generation of artificial, abnormal samples using certain techniques [35] can dramatically improve the accuracy of the classification.

The SVM algorithm has proven to be efficient when applied to SHM data. This tool is especially useful because of its adaptability to different types of data, including strain, temperature, and vibration. For instance, it has been used in a one-vs.-one setting for the assessment of paint conditions based on collected hyperspectral signatures [36]. A multiclass SVM algorithm was applied for damage detection in steel beams using critical statistical features extracted from ultrasonic wave data [37]. Sheng et al. [38] further worked on fault diagnosis in turbopump rotors using a multiclass SVM algorithm. Widodo et al. [39] used SVM classification to assess the structural health of bearings in machinery by training their classifiers using simulated acoustic emission and vibration data collected under different simulated loading conditions. Oh et al. [40] used SVM algorithms to estimate the dynamic properties of concrete strength based on ultrasonic harmonic waves collected using PZT wafers. Moreover, Park et al. [41] conducted SHM of railroad tracks using guided waves by developing an SVM classifier for damage detection and classification.

Moreover, ML has been used in many applications for pipeline structures. Chen et al. [42] developed a classifier using SVM ML to detect the NPW in a signal and differentiate it from other types of pressure drops, reaching a testing accuracy of 90%. Ying et al. [43] worked on leakage detection in pipes by collecting data from PZT sensors while pressurizing and depressurizing the pipe. Extracted features from each set of measurements were fed to an ML algorithm that was able to detect leakages with an accuracy greater than 90%. The best-used ML technique was a combination of an SVM algorithm and AdaBoost, which correctly identified damaged specimens with an accuracy of 93.5%. A leakage detection and warning system was developed by Qu et al. [44]. These authors used a distributed fiber optic vibration sensor to collect data while simulating a leakage, digging around the buried pipe, and walking in the neighborhood of the buried pipe. They managed to create an SVM classifier that was able to distinguish between those three events with an accuracy of 95%. These authors were also able to locate the event within 200 m on a 35-km pipe. Moreover, Chuang et al. [45] developed an ML algorithm using collected acoustic data from the surface above a buried water distribution pipeline. The authors trained classifiers using different inputs and different ML algorithms to detect damaged and leaking sections of the pipe. A detection accuracy of 99.4% was obtained using a Convolutional Neural Network (CNN), and a detection accuracy of 98% was obtained using an SVM classifier. Similarly, Santos et al. [46] detected pipeline leakages based on acoustic signals. These authors developed two neural networks that were respectively specialized in the detection and localization of the leakages. A localization error smaller than 2 m was achieved on a 60-m testbed.

3. Methodology

This section describes the developed methodology used in this study. The design of the testbed, the tested scenarios, and the development of the classifiers are discussed.

3.1. Sensing using FBG sensors

FBG strain sensors were used in this study to gather data. This sensing method works by sending a spectrum of light through a fiber optic cable on which an FBG is engraved. When that spectrum reaches the FBG section of the fiber, part of it is reflected to the interrogation unit, and the rest of the signal is transmitted through the fiber [47,48]. The reflected part of the spectrum is centered around a single wavelength of light called the Bragg wavelength. This wavelength can be calculated using Eq. (1), where λ_b is the reflected wavelength, n_{eff} is the effective index of refraction, and Λ is the grating period of the FBG.

$$\lambda_b = 2n_{eff}.\Lambda \tag{1}$$

Given that this reflected wavelength depends on the grating period, changing the distance between the gratings by stretching the fiber or changing its temperature results in a shift in wavelength that can be detected by the interrogation unit. The sensor response to an applied strain (ε) is given by Eq. (2).

$$d\lambda_b = 2\Lambda dn_{eff} + 2n_{eff}\mathrm{d}\Lambda \tag{2}$$

$$\frac{d\lambda_b}{\lambda_b} = \frac{dn_{eff}}{n_{eff}} + \frac{\mathrm{d}\Lambda}{\Lambda} \tag{3}$$

In Eq. (3), $\frac{dn_{eff}}{n_{eff}}$ can be written as $\{\frac{n_{eff}^2}{2}[p_{12} - \nu(p_{11} + p_{12})]\varepsilon\}$ where p_{11} and p_{12} are Pockels coefficients of the strain-optic tensor, and ν is the material's Poisson ratio. Also, $\frac{\mathrm{d}\Lambda}{\Lambda}$ can be substituted by the fiber strain ε. With further mathematical manipulations and grouping of constants, the strain detected by the FBG can be linked to the shift in wavelength by Eq. (4).

$$\frac{d\lambda_b}{\lambda_b} = K_\varepsilon \varepsilon \tag{4}$$

In this equation, $K_\varepsilon = 1 - \frac{n_{eff}^2}{2}[p_{12} - \nu(p_{11} + p_{12})]$. Given that K_ε is the only constant in this equation, it defines the strain sensitivity of the FBG.

For temperature sensing (T), similar derivations can be performed by taking into consideration the thermal-optical tensor ζ and the fiber's thermal expansion coefficient α.

$$d\lambda_b = 2\left(\Lambda\frac{dn_{eff}}{dT} + n_{eff}\frac{\mathrm{d}\Lambda}{dT}\right)dT \tag{5}$$

$$\frac{d\lambda_b}{\lambda_b} = \left(\frac{dn_{eff}}{n_{eff}dT} + \frac{1}{\Lambda}\frac{\mathrm{d}\Lambda}{dT}\right)dT \tag{6}$$

$$\frac{d\lambda_b}{\lambda_b} = (\zeta + \alpha)dT \tag{7}$$

$$\frac{d\lambda_b}{\lambda_b} = K_t dT \tag{8}$$

The temperature of the FBG can then be linked to the shift in wavelength by Eq. (8). K_t, being the only constant in this equation, characterizes the fiber's temperature sensitivity.

In this work, FBG sensors were used as strain sensors, mounted directly onto the pipe surface. Some of the sensors were regular FBG strain sensors, whereas the others were temperature-compensated strain sensors. These units encase two FBGs, one of which is tightly connected to the frame and is therefore used to measure strain, and the other is mounted onto a different piece of metal that is insusceptible to changes in strain and is therefore used to measure temperature. By using these two wavelength values as well as the host material's modulus of elasticity and the thermal expansion coefficient, the interrogation unit can calculate the temperature-compensated strain of the sensors.

3.2. Experimental approach

Experiments were conducted by recording the strain data of the FBG sensors in multiple scenarios with different pressures, flow

rates, leakage locations, and hole leakage sizes. The main concern was to replicate possible leakage scenarios in actual transportation pipelines. The sensors were placed on the pipeline in different locations and orientations (longitudinal and circumferential) to study the effects of these factors on the collected measurements and, consequently, the obtained results. The data recording was performed for all the sensors simultaneously for each experiment, eliminating any variability in the parameters that may occur between experiments.

All collected data was gathered from the interrogation of the FBG sensors. As previously mentioned, by attaching the fibers directly to the pipe's surface, the measured fiber strain and its variation were expected to reflect the properties of the flow inside the pipe. By performing tests under different flow and leakage conditions, the effects that these parameters have on the output signals can be determined.

By comparing the signals obtained in various scenarios (by changing one parameter at a time), it was possible to observe the effects that a specific parameter had on the readings.

3.3. SVM ML implementation

Classifications were performed for pressure, flow rate, leakage location, and leakage size using an SVM algorithm. The choice of the ML algorithm was based on preliminary tests that were performed using different approaches, including deep learning methods such as a CNN [49] and an SVM algorithms, which revealed that the later classifiers performed well on the gathered strain data.

An SVM algorithm can be used for binary, multiclass, and regression classification problems [50–52]. SVM algorithms works by separating the initially given data (training data) by hyperplanes and decision surfaces with a maximal margin. When points are not linearly separable, they are mapped into a higher-dimensional feature space using different possible kernel functions, such as polynomial, Gaussian, and radial basis, among others. Initially, SVM classifiers were only capable of solving binary problems. When the problem consists of differentiating between more than two types of data points, the implementation of binary transformation allows the use of an SVM algorithm for multiclass problems. The one-vs.-one binary transformation works by separating the data belonging to two classes at a time, building hyperplanes and decision surfaces for each pair of classes. The one-vs.-all transformation, however, works by separating the data belonging to each class from the data belonging to all other classes [53]. These two transformations were tested in this study. An illustration of the differences between these transformations is shown in Fig. 1.

In the case of linearly separable problems, consider a training set of N samples $(\boldsymbol{x}_i, y_i)_{i=1}^{N} \in R^n$, where x_i is the input vector labeled by $y_i \in [0, 1]$. An SVM algorithm seeks to separate the binary data by finding the optimal hyperplane that can maximize the distance between the binary points belonging to different classes [30]. The hyperplane is defined by parameters $\boldsymbol{\omega}$ and b with the equation $\boldsymbol{\omega}.\boldsymbol{x} + b = 0$. The maximum margin hyperplane can be identified as the solution of the following minimization problem:

$$z = \frac{1}{2}\boldsymbol{\omega}.\boldsymbol{\omega} + C\sum_{i=1}^{N}\xi_i \tag{9}$$

$$s.t y_i(\boldsymbol{\omega}\boldsymbol{x}_i + b) \geq 1 - \xi_i, i = 1, \cdots, N \tag{10}$$

where the $C\sum_{i=1}^{N}\xi_i$ component is used to control the number of misclassified points, and$C$is the penalty parameter.

To predict the class of a new point $\boldsymbol{x}$, the decision function is computed as follows:

$$f(x) = sign(\sum_{i=1}^{N} y_i\alpha_i < \boldsymbol{x}_i, \boldsymbol{x} > + b) \tag{11}$$

Thus, all the training points have $\alpha = 0$ except for the support vectors, which have a relatively small percentage of the entire training points, or $\boldsymbol{x}_i$. Therefore, to make a prediction for the new point $\boldsymbol{x}$, the inner product between $x$and the support vectors must be calculated.

SVM classifications are performed on the processed signals using MATLAB. To choose the most adequate and efficient learning settings, preliminary tests were performed using different combinations of kernels (linear, second-degree polynomial, and Gaussian), learning approaches (one-vs.-one and one-vs.-all), and standardization. These tests revealed that the best learning settings for this application are no standardization and a Gaussian kernel. For flow rate classifications, a one-vs.-one binary transfor-

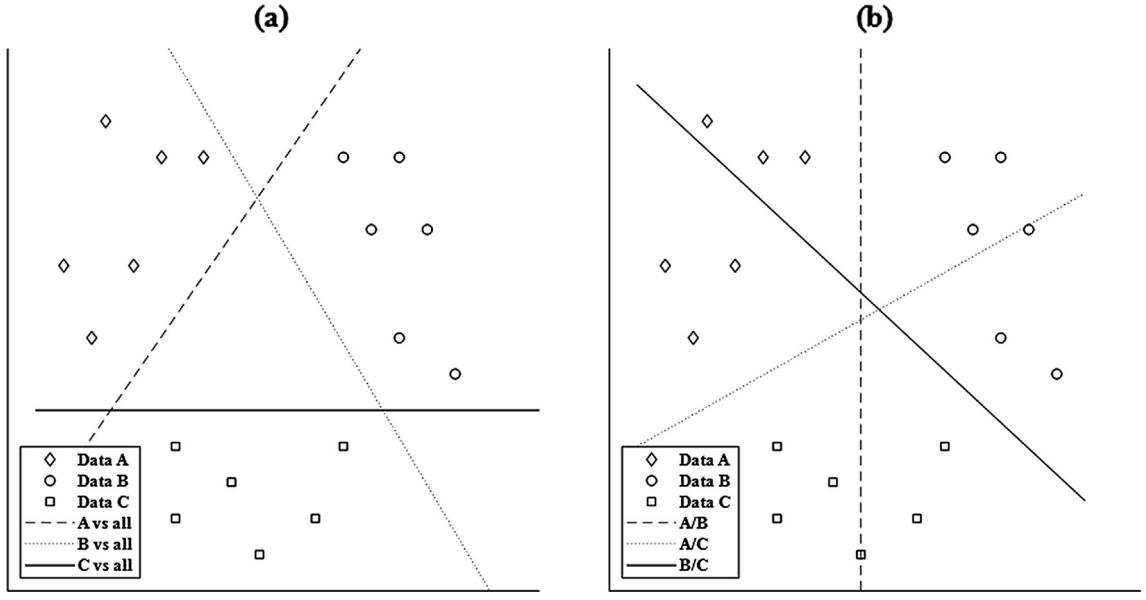

Fig. 1. Example showing the differences between (a) one-vs.-one and (b) one-vs.-all binary SVM transformations.

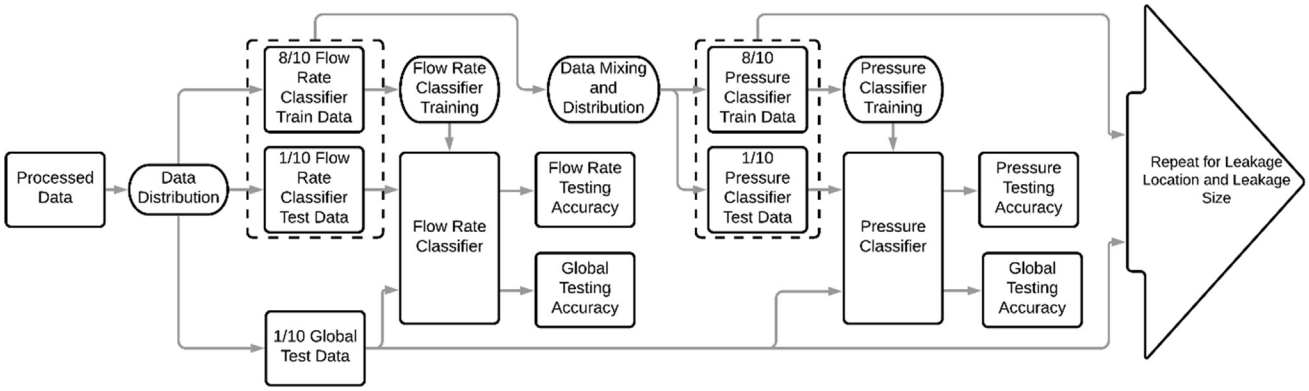

Fig. 2. The sequence for training and testing the machine learning classifiers.

mation is preferable, whereas a one-vs.-all transformation is more accurate for the classifications of pressure and leakage location. When it comes to leakage size, the results were equally good using both approaches.

The classification of data was approached in two ways: (a) a classifier was established and evaluated for each parameter i.e. pressure, flow rate, leakage size, and location and (b) classification of the testing data was performed using a combination of the classifiers [from (a)] for all parameters (labels) combined.

The sensor data obtained in each experiment were concatenated into a $1 \times 120{,}000$ array. The arrays corresponding to each experiment were then concatenated to form the final $720 \times 120{,}000$ array. When testing the behavior of different sensor combinations, the input data for training and testing were modified so that they only contained the concatenation of the signals of the required sensors. The distribution of data between the training and testing sets was randomized.

The data were randomly divided into a portion consisting of 90% of the data for training and testing individual parameters, and a global testing portion composed of 10% of the data for the combined labels.

The classifier of individual parameters was established by randomly subdividing the 90% data set into 80% for training and 10% for testing. When another classifier needed to be developed, the training and testing portions were randomly selected from the 90% data set, excluding the initial 10% of the data set placed aside for global testing. The pressure, flow rate, leakage location, and leakage size classifiers were each tested using the redistributed 10% of the data and the global testing data. This procedure is described in Fig. 2. The complete process was repeated 10 times to ensure the repeatability of the results.

ML was also used as a tool to optimize the number of sensors and the placement of sensors on the structures. In this work, given the small size of the testbed and the small number of sensors, all the possible combinations were tested; however, for larger sensor networks that are mounted onto longer pipes, some optimization techniques should be considered [54,55].

4. Experimental set-up

4.1. Testbed design

A testbed was specifically designed to demonstrate the feasibility of the application of FBG sensors in monitoring the operational and structural conditions of the pipelines. The testbed consists of a 5.95-m-long, black steel pipe with a nominal diameter of 10″ and a thickness of 5 mm, resulting in a total inner volume of 0.2782 m^3. The full experimental set-up is shown in Fig. 3.

Multiple design components like valves and pressure gauges must be mounted on the pipe; hence, the choice of material and wall thickness is critical to allow the welding of threadolets onto the surface for the attachment of the fittings. To facilitate the manipulation and transportation of the pipe, it was initially cut into four segments, which were reattached using flanges that were welded onto the extremities and connected with M10 bolts and nuts. Rubber gaskets were placed between the flanges to make this assembly watertight.

Six holes with different purposes were drilled into the pipe in the radial direction on the pipe. Two holes served as the inlet and exhaust of fluid through the pipe, while the others were used to simulate leakage. At the inlet of the pipe, a ball valve, pressure gauge, flow meter and pressure-reducing valve were installed. At the exit hole, a ball valve was the only needed fitting for flow control. The holes needed for leakage simulation were each equipped with a ball valve as well.

When a leak needed to be simulated from the different leakage points, an extra ball valve was placed over that point. The purpose of this addition was to allow the experiments to be replicated at specific leakage flow rates. The valve connected directly to the pipe therefore acted as an on/off switch for the leak, and the additional valve on top of it could be set to a specific opening size.

All necessary connections were established using 1-inch nominal diameter pipes, elbows, and T connections. Two 1-inch flexible connections were also used to connect the inlet to the pump and the exhaust to the tank, which also acted as a sump.

As for the instrumentation, two different types of sensors were used in the set-up; these were supplied by Micron Optics. The sensors placed in the longitudinal direction were temperature-compensated FBG strain sensors (os3155). They were spot-welded onto the pipe surface, as their design included spot-welding locations. The sensors mounted in the circumferential direction were optical strain gages (os3100) that were epoxied to the pipe's surface. The six mounted sensors were named S1 through S6, starting from the sensor closest to the pipe's inlet. Sensors S1, S3, S4, and S6 were mounted in the longitudinal direction, whereas sensors S2 and S5 were mounted in the circumferential direction. Data were collected using the Micron Optics si255 Hyperion interrogator at a rate of 1 kHz.

4.2. Tested scenarios

Experiments were performed by arranging the pipe in its normal running condition at a certain pressure and flow rate. Data recording was then initiated and set to record for 1 min. At the 30-second mark, the leakage hole equipped with the second valve was opened. This process is shown in Fig. 4.

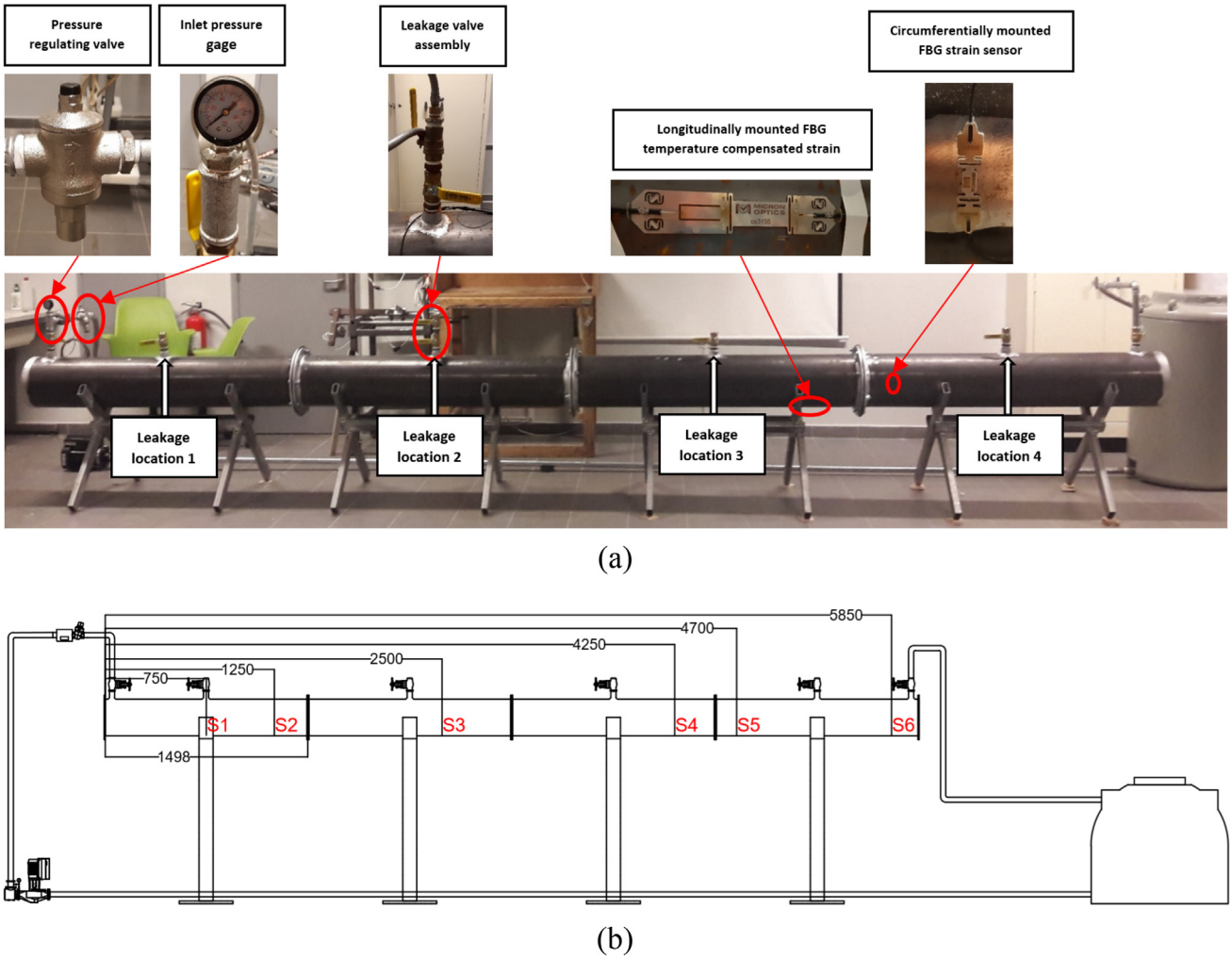

Fig. 3. Detailed layout of the (a) testbed experimental set-up and (b) testbed dimensions and sensor placement.

The collected data, therefore, corresponds to the readings obtained during 30 s of regular pipe operation under normal conditions followed by 30 s of operation with leakage under the same conditions.

Tests were performed at three different pressures (0.6, 0.8, and 1 bar) and three different flow rates (10, 15, and 20 GPM). Classifying these flow properties can then be used to assess the severity of the impact of the leakage. These variations also provided the other classifiers with a wider variety of data for learning, making them more robust and diverse. In combination with these scenarios, four leakage locations and two leakage sizes [small (~20%) and large (100%) leakage valve openings] were tested. The leakage size was controlled by specifying the opening of the secondary valve, as previously mentioned. The variations of this parameter further increased the variability of the data and ensured that the other classifiers can perform reliably, regardless of the leakage size.

The possible combinations of these parameters are shown in Fig. 5. Tests were performed on each of these combinations and repeated 10 times. The total number of tested scenarios is therefore 72, and the total number of readings is 720.

4.3. Sensor data

The data recorded by the interrogation unit contained noise that masked the signal. The signal-to-noise ratio (SNR) ranged between 1.85 and 10.23, depending on flow and leakage conditions. The data were therefore processed using MATLAB to reduce the noise. A moving mean of 400 ms was first applied for the smoothing of the data followed by a low-pass filter with a factor of 10^{-5}. The signal coming from each sensor was then zeroed and trimmed to 15 s to eliminate redundant data. This sequence is shown in Fig. 6.

The changes in surface strain measured in this study can be compared to results reported in previous studies. For instance, tests performed at a pressure of 120 kPa (1.2 bar) yielded a change of 5 to 6 με on the surface of the pipe [56]. Other tests performed at a much higher pressure of 1.2 MPa (12 bar) inflicted 90–120 με on the surface of the pipe [57]. These tests were both performed on sections of pipe similar to the one used in this work. Oil transportation pipelines typically operate at much higher pressures than those used in the testbed in this study [58], and hence, at a much higher hoop stress. Therefore, in the presence of leakage, these high pressures will increase the observed strain drops due to the higher-pressure amplitude of the NPW, guaranteeing a higher SNR and more easily observable and clear strain drop/ patterns.

5. Results and analysis

In the following section, the effects of pressure and flow variation, as well as leakage size and location, on the collected strain data are scrutinized. Each of these parameters was studied and analyzed. In this section, only data from sensors 1 and 2, which

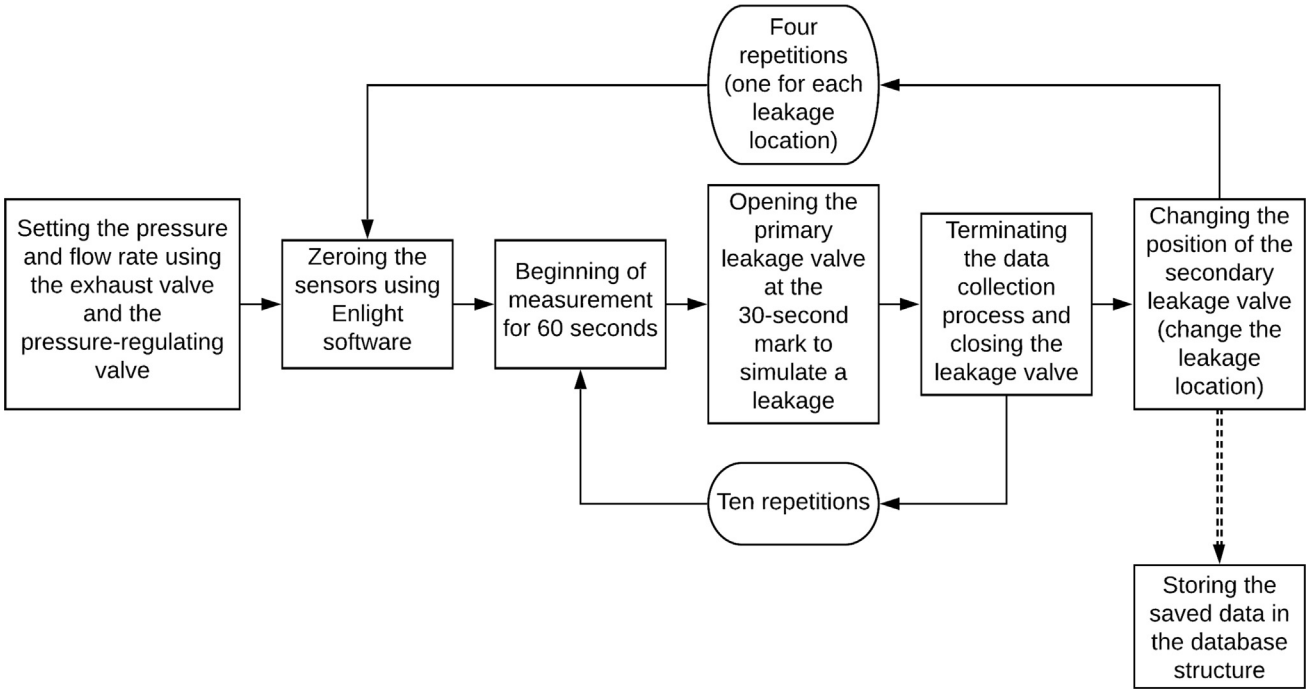

Fig. 4. Experimental procedure flow chart.

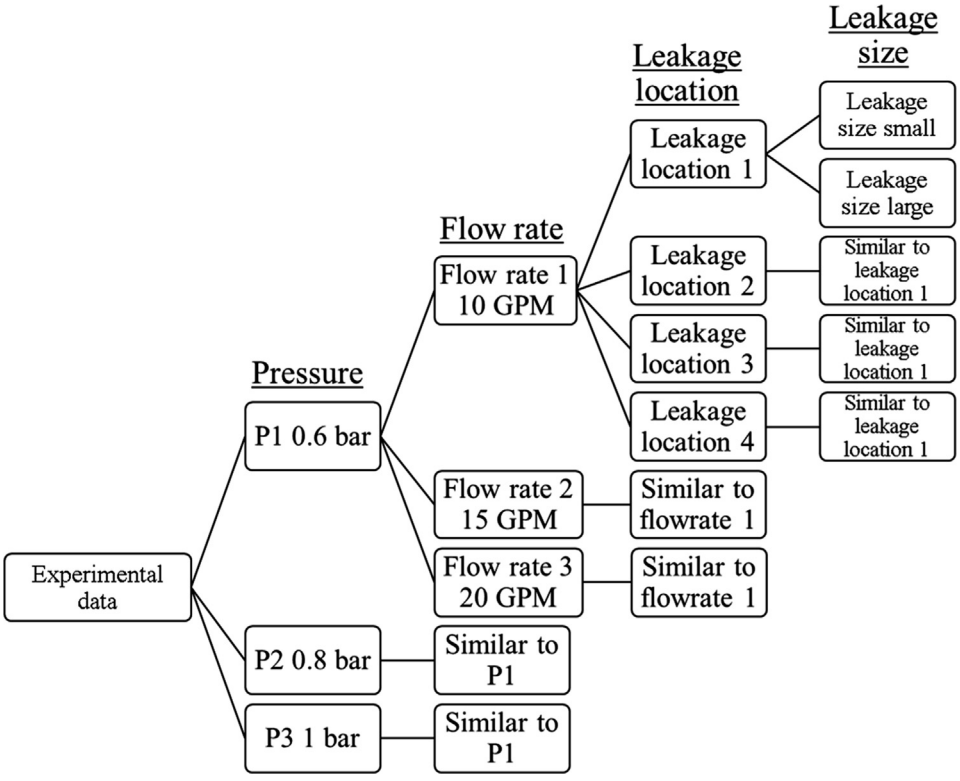

Fig. 5. The various tested scenarios.

were placed 75 cm and 125 cm from the pipe extremity on the inlet side, are presented. It is important to note that any sensor combination yields similar results. The choice of these two sensors was specifically made as they best display the patterns (with maximal variation in the amplitude). Also, these sensors were placed on the pipe in different directions, showing the effect of sensor orientation on the results (sensor 1 was aligned longitudinally; sensor 2 was aligned circumferentially). Moreover, this section contains the ML results.

The gathered data contains signals obtained from experiments performed with all the possible combinations of leakage and flow properties. To study the effects of a change in any single parameter

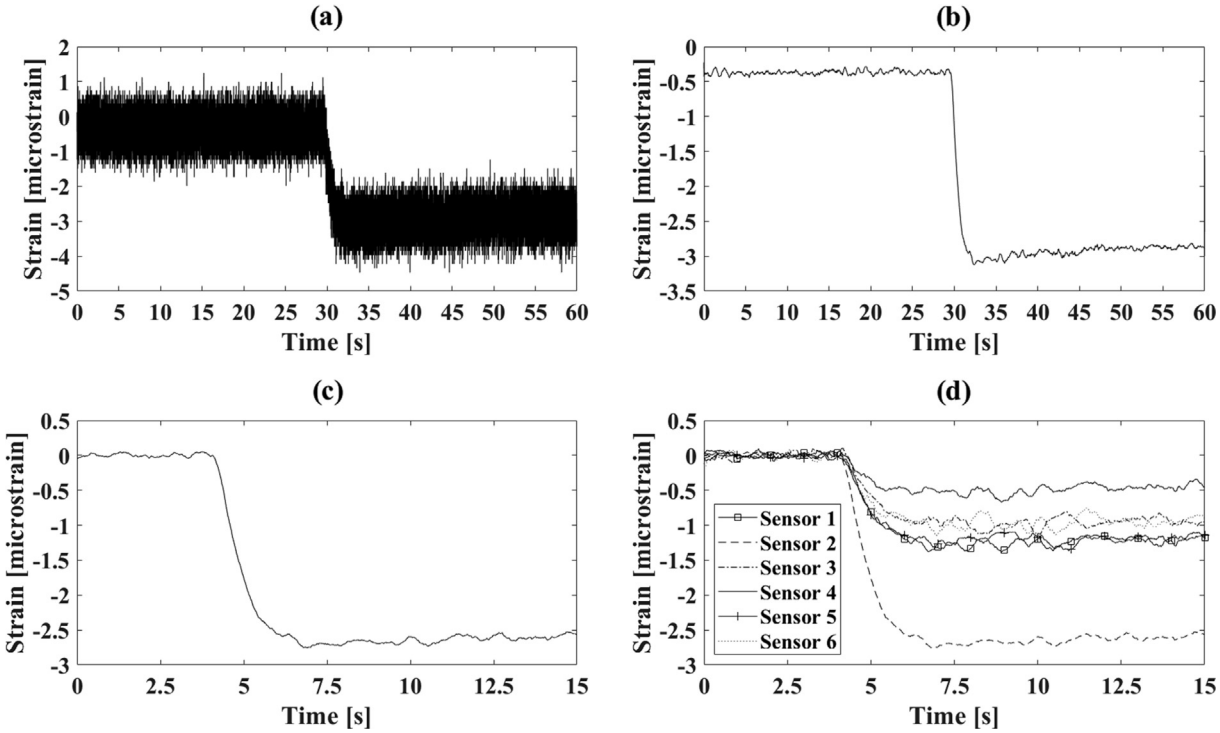

Fig. 6. Steps followed for signal preparation: (a) raw data from one sensor (S1), (b) signal after noise reduction, (c) trimmed and zeroed signal, and (d) stacked signals of six sensors.

on the shape of the signal, the data were grouped based on that parameter. For instance, to study the effects of a change in pressure, the signals gathered at each value (0.6, 0.8, and 1 bar) were grouped. These data however contained signals obtained at different flow rates, leakage locations, and leakage sizes. Therefore, to specifically observe the effects of changes in pressure, multiple comparisons had to be made to ensure that those observations were true, regardless of the change in the other parameters. For example, one plot compared signals gathered at a 10-GPM flow rate and a large leakage at varying pressures, and another compared signals gathered at a 20-GPM flow rate and a large leakage at those same pressure variations. If the observations made in these two cases are similar, the conclusion that this observation is independent of flow rate becomes reliable.

5.1. Effect of flow rate variation on strain measurement

To observe the effects of flow rate on the measured strain, the values measured for different flow rates in four different scenarios are presented. The different scenarios also provide insight into how different parameters, such as leakage location and leakage hole size, can change the observed effects of flow rate variations.

The first scenario consists of measurements taken at a pressure of 0.8 bar for a small leakage size and leakage location (L1) (closest to the inlet valve and the presented sensors). Fig. 7 shows a decrease in the strain level that resulted from the increase in flow rate. Both circumferential and longitudinal sensors were able to detect this decrease in strain level. A drop of approximately 3% was noticed between flow rates of 10 and 15 GPM and a drop of approximately 10% between flow rates of 15 and 20 GPM.

The second scenario consists of measurements taken at the same pressure (0.8 bar) but with a change in the leakage position to L4 (farthest point from the inlet). Fig. 8 shows the decrease in strain drop that resulted from the increase in flow rate. In this case, where the leakage is far from the sensors, both sensors detected a drop of approximately 4% between flow rates of 10 and 15 GPM and a drop of approximately 9% between flow rates of 15 and 20 GPM. The results shown in Table 1 reveal that the data are very similar to those observed in the previous scenario, suggesting that leakage location has a minimal or negligible effect on strain amplitude.

The third scenario is similar to the first scenario but with an increase in the leakage size, that is, a pressure of 0.8 bar with a large leakage size at leakage location L1. By comparing this scenario to the first, information about the effects of leakage size on the measured strain can be obtained. Fig. 9 shows the effects of a change in flow rate on strain measurements when a large leakage occurs. In this case, the strain drop is larger than that resulting from a smaller leakage. For any given flow rate and any given sensor, the observed strain drop is increased by more than 40%. When comparing flow rates, the sensors detected changes that differed from the ones observed for a smaller leakage. When shifting from a 10-GPM to a 15-GPM flow, longitudinal sensors were shown to detect a 10% decrease in strain drop, whereas circumferential sensors detected a 3.5% decrease. However, when shifting from a 15-GPM to a 20-GPM flow, the longitudinal sensor strain drop decreased by 5.5%, and the circumferential sensor strain drop decreased by 8%.

The fourth scenario consists of measurements taken at a pressure of 0.8 bar with a large leakage size at leakage location L4. The observations made in this scenario will help determine whether the leakage location affects the strain drops differently when a large leakage occurs. Fig. 10 shows the effects of a change in the flow rate on the strain measurements when a large leakage occurs far from the sensor position.

The observation that was made in the third scenario (i.e., a larger strain drop relative to cases with a small leakage) is still observable with a large leakage size at a distant leakage position. The strain drop variation, however, is minimal in this case and varies by a maximum of 3.5%. The drop values obtained at a 15-GPM

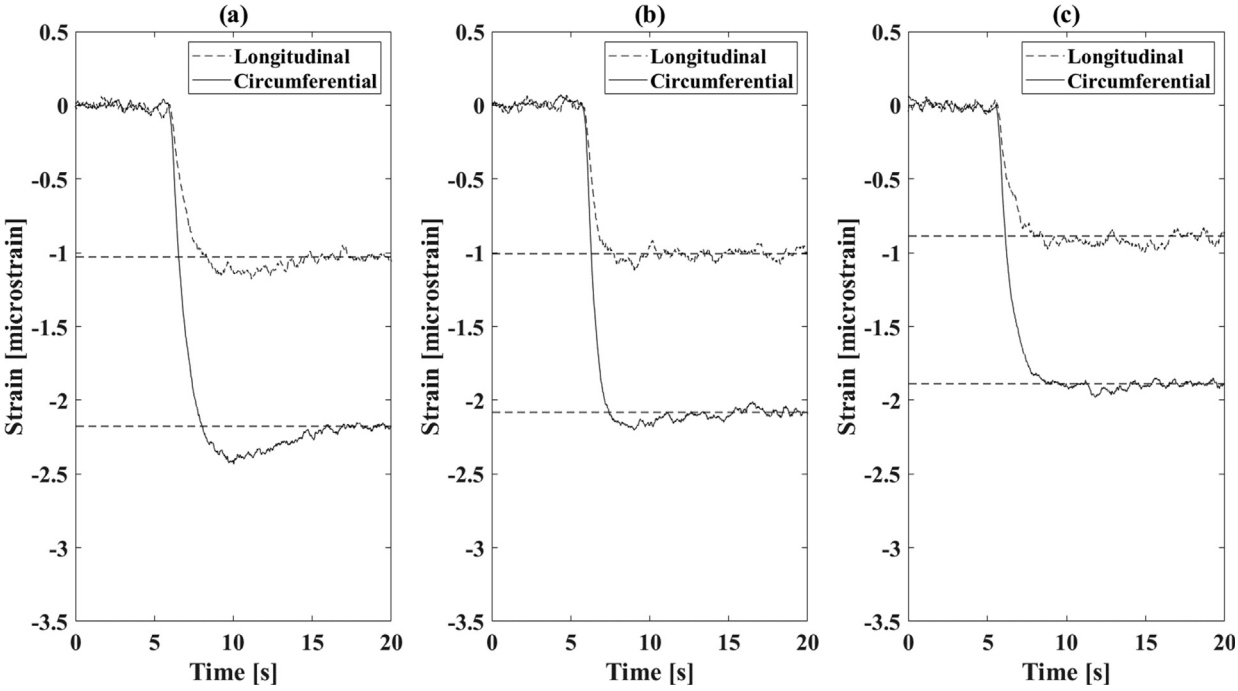

Fig. 7. Flow rate comparisons at (a) 10, (b) 15, and (c) 20 GPM at 0.8 bar at leakage location L1 with a small leakage size.

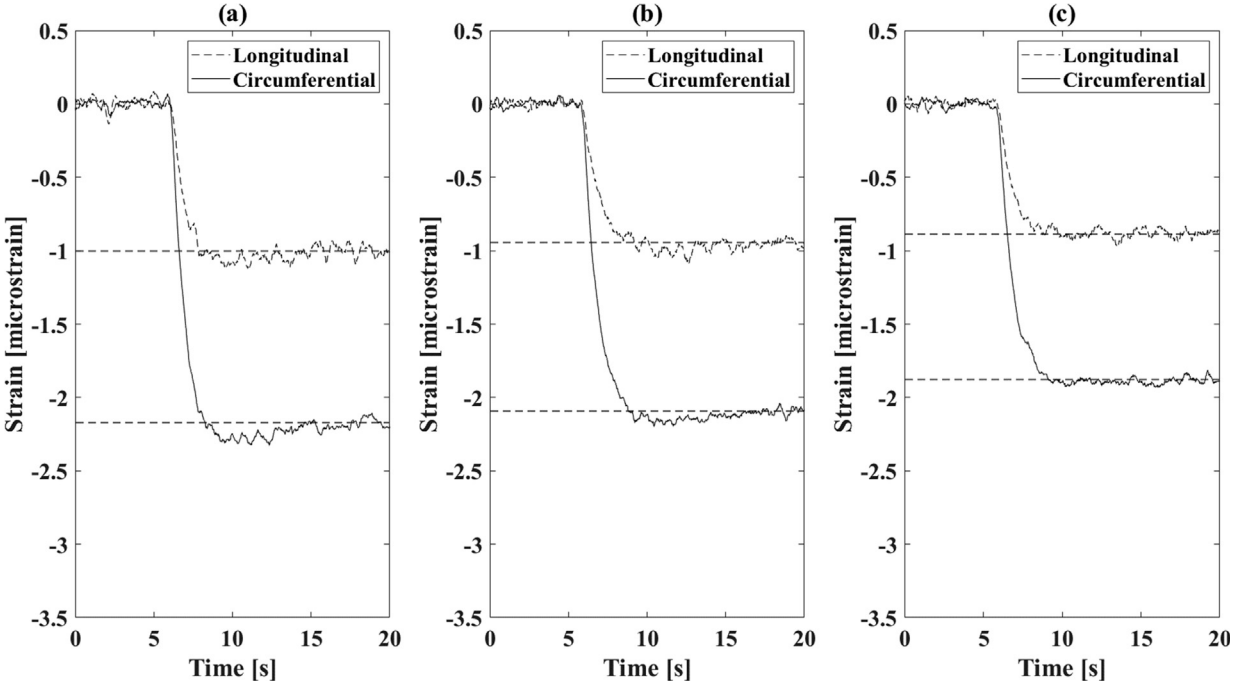

Fig. 8. Flow rate comparisons at (a) 10, (b) 15, and (c) 20 GPM at 0.8 bar at leakage location L4 with a small leakage size.

flow rate at both leakage locations are similar, varying by only 0.01 microstrain, suggesting that flow rate in a pipe with a large leakage at a distant location, relative to the size of the pipe under study, has a smaller effect on the observed strain drop.

5.2. Effects of pressure variations on strain measurements

To observe the effects of pressure on the measured strain, the values measured for different pressures in four different scenarios are presented. Similar to data presented earlier, these scenarios will allow analyses of the effects of pressure variations under different flow and leakage conditions.

The first scenario shows the effects of different pressures when a small leakage occurs at a location that is close to the sensors (L1). The main difference in the measured strain between pressure and flow rate variations is that an increase in flow rate results in a smaller strain drop, whereas an increase in pressure results in a larger strain drop. Fig. 11 shows the increase in strain drop that results from the increase in pressure. The longitudinal and circumferential sensors detected increases of 31% and 18%, respectively, in

Table 1
Strain values recorded for different flow rate variation scenarios.

	10 GPM	15 GPM	20 GPM
Strain in με for flow rate variation at 0.8 bar at leakage location L1 with a small leakage size			
Longitudinal	−1.03	−1.01	−0.89
Circumferential	−2.18	−2.08	−1.89
Strain in με for flow rate variation at 0.8 bar at leakage location L4 with a small leakage size			
Longitudinal	−1	−0.95	−0.89
Circumferential	−2.17	−2.09	−1.88
Strain in με for flow rate variation at 0.8 bar at leakage location L1 with a large leakage size			
Longitudinal	−1.84	−1.66	−1.57
Circumferential	−3.08	−2.98	−2.74
Strain in με for flow rate variation at 0.8 bar at leakage location L4 with a large leakage size			
Longitudinal	−1.71	−1.65	−1.64
Circumferential	−3.01	−2.97	−2.87

strain drop when the pressure was changed from 0.6 to 0.8 bar (33% increase in pressure). Both sensors also detected a 13.5% average increase in strain drop when the pressure was further increased from 0.8 to 1 bar (25% increase in pressure).

The second scenario consists of measurements taken at a flow rate of 15 GPM with a small leakage size at a distant leakage location (L4). When the leakage location is far from the sensors, the pressure variation seems to have a slightly smaller effect on their measured strain. These effects are presented in Table 2, which shows the similarity between the values recorded for the first two scenarios. The longitudinal and circumferential sensors detected increases of 15% and a 13%, respectively, in strain drop when the pressure was changed from 0.6 to 0.8 bar. When changing the pressure from 0.8 to 1 bar, these values changed to 18% for the longitudinal sensor and 12% for the circumferential sensor. These patterns can be seen in Fig. 12.

The third scenario consists of measurements taken at a flow rate of 15 GPM with a large leakage size at leakage location L1. This scenario serves as an indicator of the effects of changes in leakage size compared to the first scenario. Similar to the case shown in Fig. 9, a larger leakage size increased the strain drop relative to a small leakage size. The larger leakage size also accentuates the effects of pressure variations. This fact becomes evident when comparing Fig. 11 and Fig. 13. The longitudinal sensor's measured strain increased by 32% and 15% when the pressure changed from 0.6 to 0.8 bar and from 0.8 to 1 bar, respectively. The measured strain of the circumferential sensor increased by 24% over each pressure increment.

The fourth scenario consists of measurements taken at a flow rate of 15 GPM with a large leakage size at leakage location L4. The strain values obtained with a large leakage at a far leakage location are similar to those obtained at a close leakage location with the same leakage size, as seen by comparing Figs. 13 and 14. The strain values varied by a maximum of 0.05 microstrain, indicating that the leakage position has little to no effect on the strain drop (given the size of the testbed under study).

5.3. Effect of leakage size variation on strain measurement

As inferred from the variations of pressure and flow rate, the smallest strain drop occurs when the pipe flow is at a small pressure and a high flow rate. The largest strain drop occurs in the opposite configuration. Comparing the measured strain in these two cases for different leakage sizes can provide insight into the extent of the effects of leakage size.

As shown in Fig. 15, using longitudinal sensors, the strain drop difference between a large and a small leakage is 70% when comparing maximal drops (a and c) and 64% when comparing minimal drops (b and d). Circumferential sensors, however, detect a 41% change in both scenarios. This finding indicates that longitudinal sensors may be more sensitive to changes in leakage size than circumferential sensors.

5.4. Effect of leakage location on strain measurement

Although leakage location does not affect the magnitude of the strain drop, it does affect the shape of the strain curve. To observe

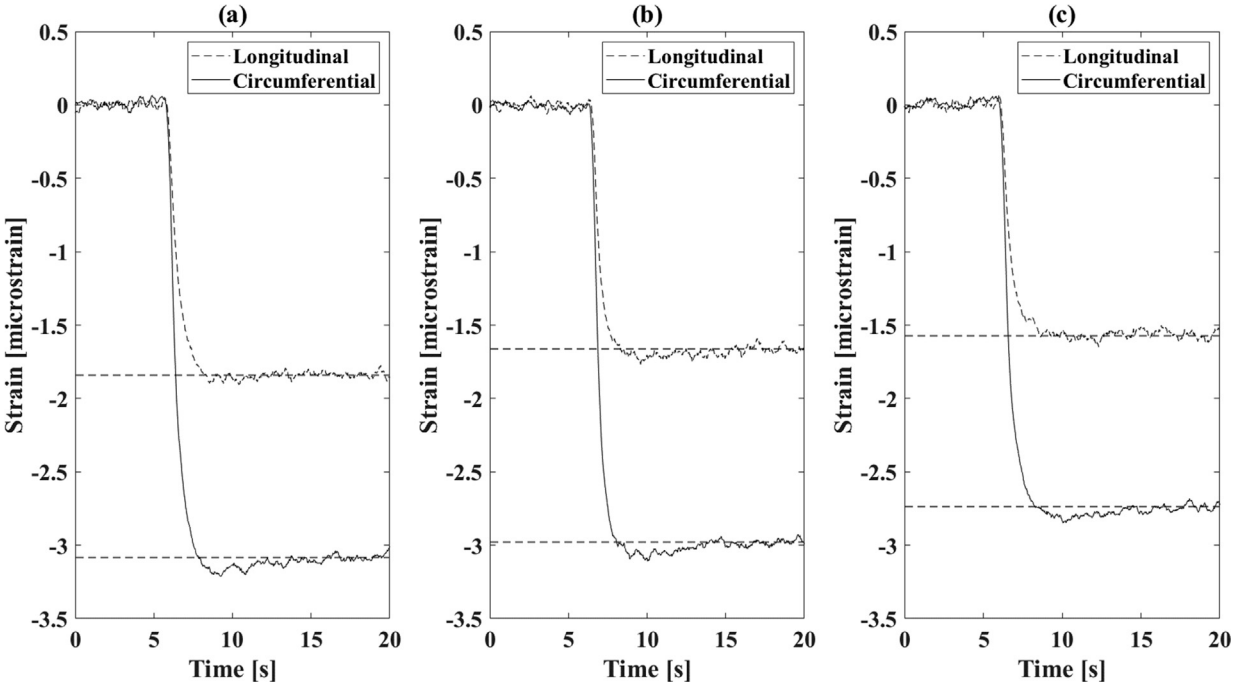

Fig. 9. Flow rate comparisons at (a) 10, (b) 15, and (c) 20 GPM at 0.8 bar at leakage location L1 with a large leakage size.

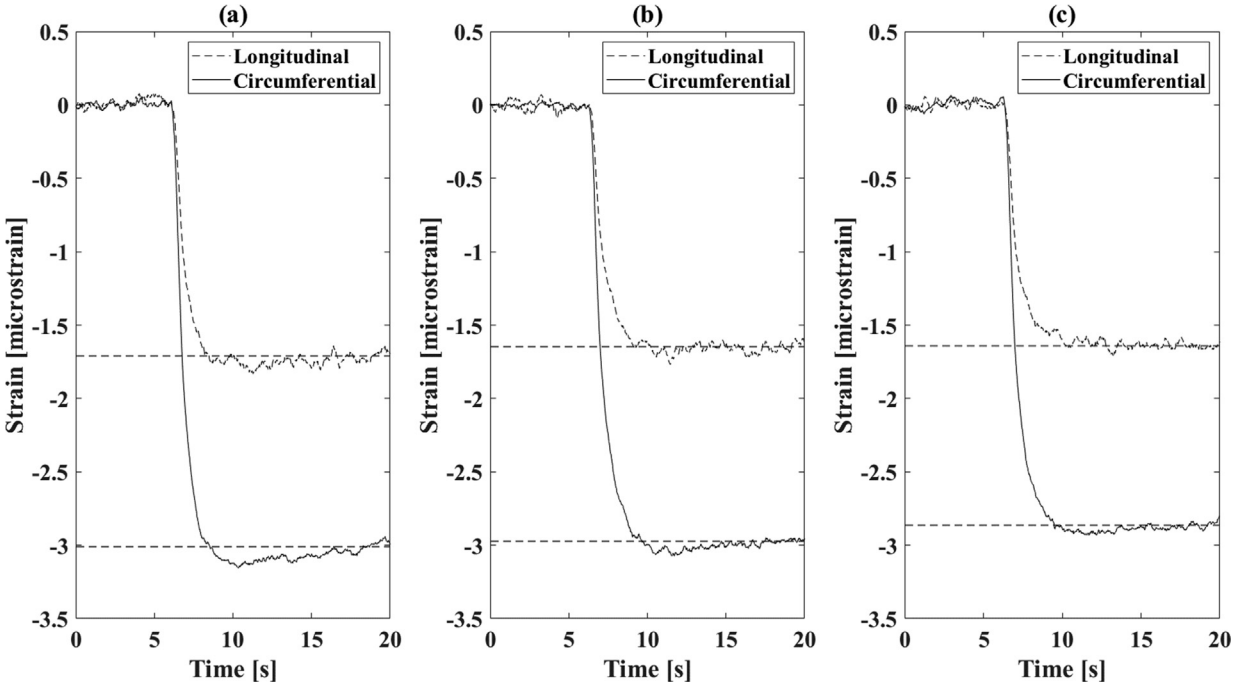

Fig. 10. Flow rate comparisons at (a) 10, (b) 15, and (c) 20 GPM at 0.8 bar at leakage location L4 with a large leakage size.

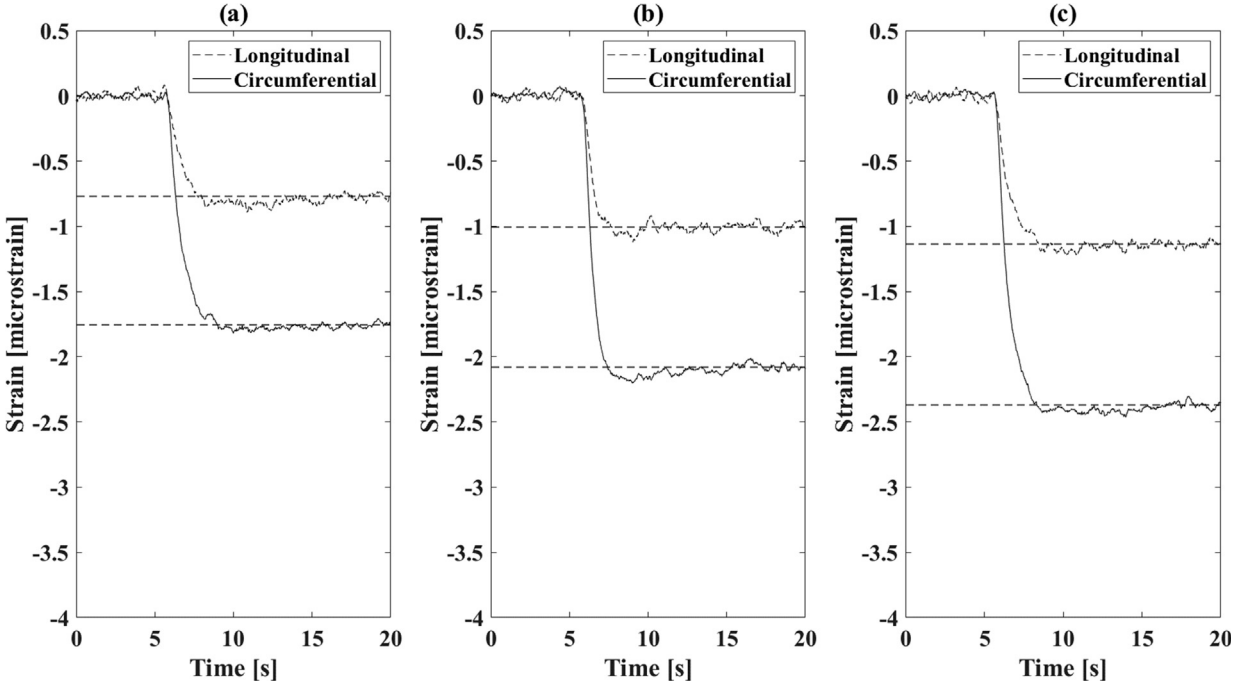

Fig. 11. Pressure comparisons at (a) 0.6, (b) 0.8, and (c) 1 bar at 15 GPM at leakage location L1 with a small leakage size.

this effect, leakage location was varied in different scenarios. To observe this change, the scenarios in question are the ones that yielded the largest and smallest strain drops using both large and small leakage sizes. The most easily observable difference between the signals obtained by changing leakage location was the gap between the intercepts of sensors 1 and 2, using sensor 1′s final steady-state strain value. This gap is highlighted in Fig. 16.

The time difference between the two intercepts appears to increase the more distant the leakage location is relative to the sensor's position. With few exceptions, this pattern holds for different flow rates, pressures, and leakage sizes. These data are presented in Table 3.

5.5. ML results

The classifications of pressure, flow rate, leakage size, and leakage location were performed using different sensor combinations to study the effects of sensor position and orientation on the

Table 2
Strain values recorded for different pressure variation scenarios.

	0.6 bar	0.8 bar	1 bar
Strain in με for pressure variations at 15 GPM at leakage location L1 with a small leakage size			
Longitudinal	−0.77	−1.01	−1.14
Circumferential	−1.76	−2.08	−2.37
Strain in με for pressure variations at 15 GPM at leakage location L4 with a small leakage size			
Longitudinal	−0.84	−0.97	−1.10
Circumferential	−1.80	−2.13	−2.39
Strain in με for pressure variations at 15 GPM at leakage location L1 with a large leakage size			
Longitudinal	−1.31	−1.73	−2.00
Circumferential	−2.38	−0.97	−-3.69
Strain in με for pressure variations at 15 GPM at leakage location L4 with a large leakage size			
Longitudinal	−1.30	−1.68	−1.96
Circumferential	−2.42	−3.02	−3.64

results. Classifiers for each flow and leakage property were developed by training and testing multiple times on different sets of data. In this section, the accuracy of these classifiers and an analysis of senor configuration are presented.

Classifiers were developed using all possible sensor combinations. With the six sensors present on the testbed, 63 different configurations were possible. Also, one-vs.-one and one-vs.-all learning approaches were established, resulting in 126 different configurations for each of the four classifiers.

5.5.1. Flow rate classification

As suggested earlier, flow rate variations result in changes in the maximal strain drop. However, these changes may be caused by a variation in pressures, flow rates, and leakage sizes.

The performance of flow rate classifiers and the choice of the number of sensors and sensor configurations that work best for flow rate classification are presented in this section.

Table 4 shows that when using certain sensor combinations, the accuracy of the prediction can be as high as 88.33%. The average testing accuracy for flow rate classification was 80%, and the lowest obtained accuracy was 57%. These findings show that classification accuracy lies on a large spectrum and requires certain sensor configurations to perform well.

The best performing sensors and the most efficient number of sensors were found by counting the number of instances where each sensor appeared or where a certain number of sensors appeared in the top 40 results (top 32% of results).

The number of sensor combinations possible changes with the number of sensors used. For example, six sensor combinations are possible with single sensors, whereas 20 combinations are possible with three sensors. The number of appearances was thus normalized, as seen in Table 6.

Table 5 shows that flow rate classification works best when a high number of sensors is used. It also shows that combinations consisting of fewer than three sensors have a very low chance of yielding good results.

As for specific sensors, Table 6 shows that sensor 2 positively influences the classification, as it is the most abundant in the top performing 32% of classifiers. Sensor 5, however, is the least commonly found sensor in the top-ranking classifiers. These findings show that sensor orientation (both sensors 2 and 5 are mounted along the circumferential direction) does not have a direct effect on the results. This finding may be due to the fact that sensor 2 has the highest gain (high SNR), and therefore, less information was lost in its signal due to filtering.

5.5.2. Pressure classification

Similar to the case of flowrate classification, pressure classifiers were developed because of the similarities that changing pressure and flow rate created in the patterns. This classifier was therefore used as a tool to differentiate between such cases.

The performance of pressure classifiers and the choice of the number of sensors and sensor configurations that work best for pressure classification are presented in this section.

Table 7 shows the 10 best ranking pressure classifiers based on the testing accuracy. As shown in the table, pressure classification can be performed with high accuracy. The best ranking classifier

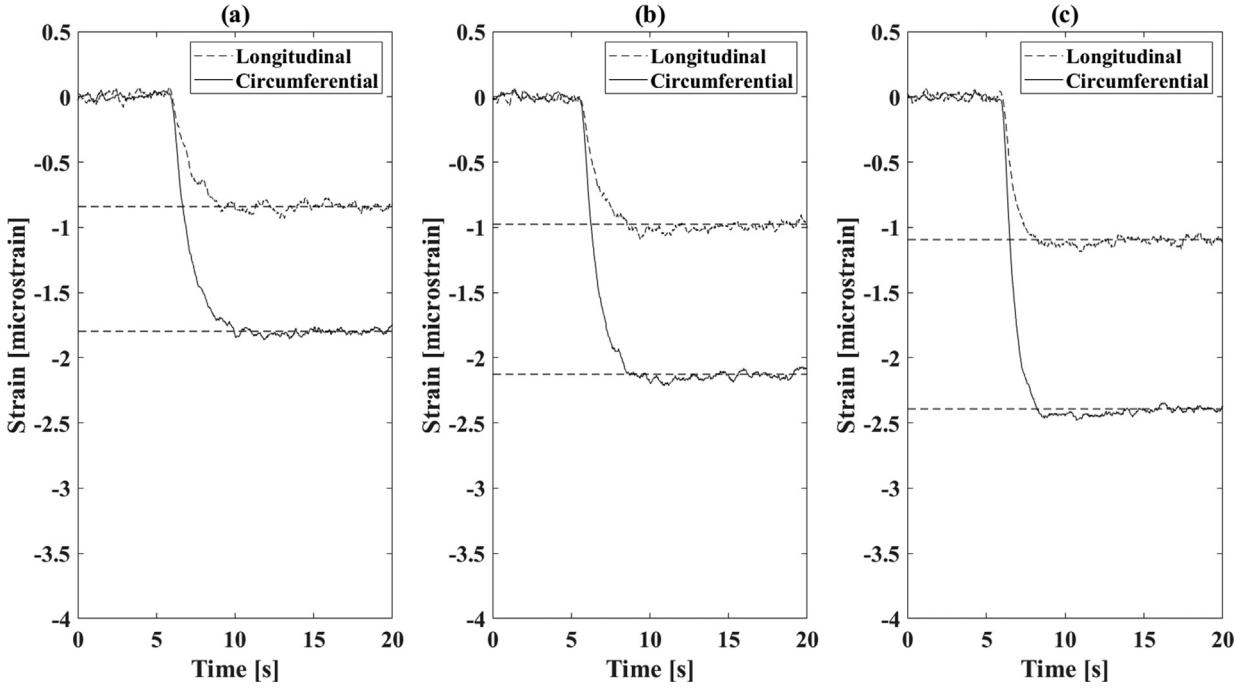

Fig. 12. Pressure comparisons at (a) 0.6, (b) 0.8, and (c) 1 bar at 15 GPM at leakage location L4 with a small leakage size.

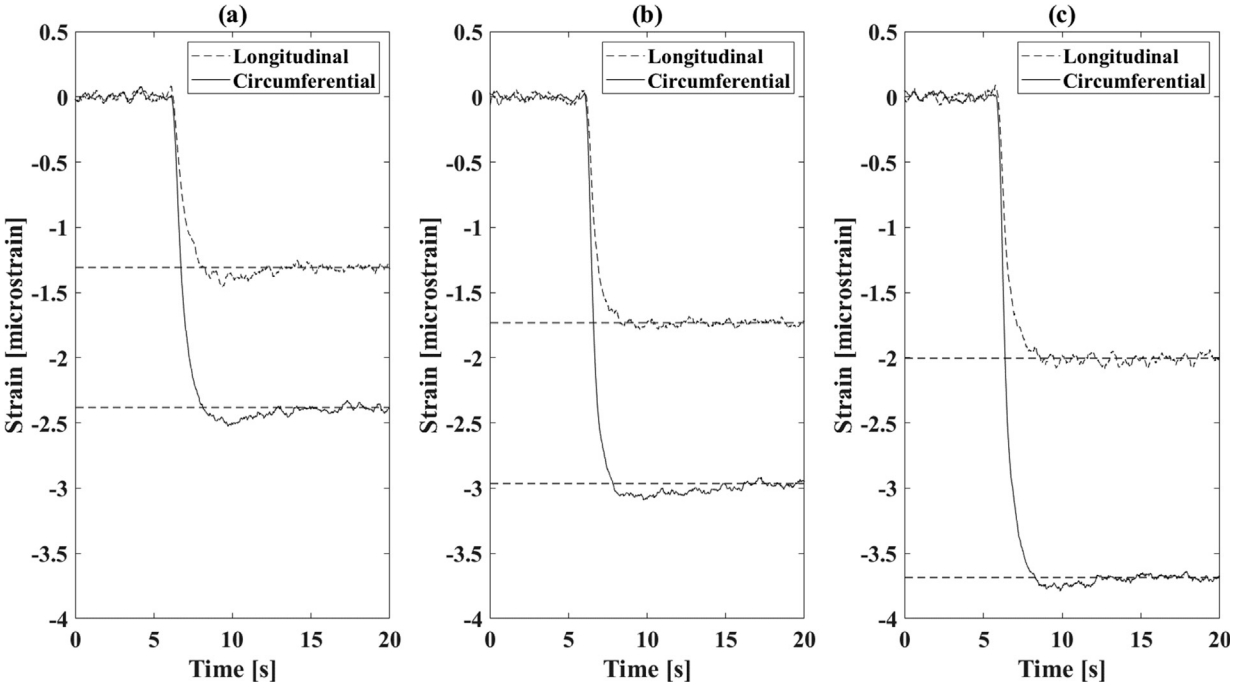

Fig. 13. Pressure comparisons at (a) 0.6, (b) 0.8, and (c) 1 bar at 15 GPM at leakage location L1 with a large leakage size.

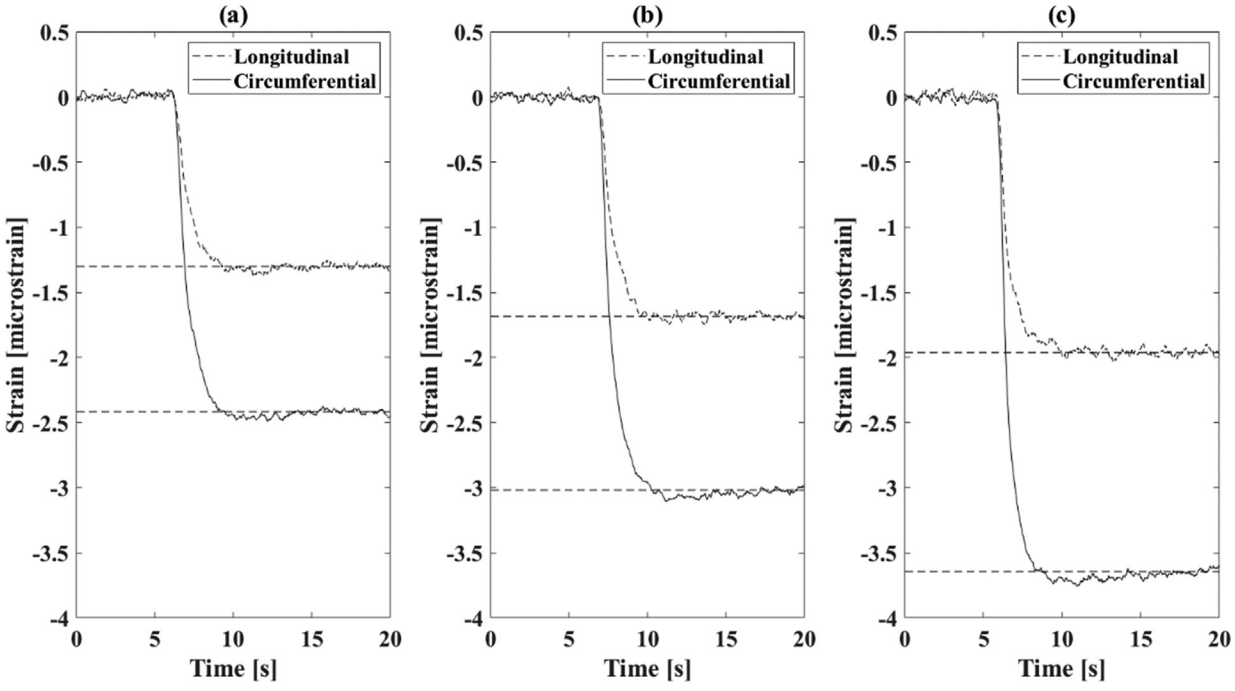

Fig. 14. Pressure comparisons at (a) 0.6, (b) 0.8, and (c) 1 bar at 15 GPM at leakage location L4 with a large leakage size.

was trained and tested using the data from sensors 1, 2, 3, and 5, achieving a 96.92% training accuracy and a 95% testing accuracy. The average testing accuracy for pressure classification was 90%, indicating that pressure classification can be performed efficiently with a wide variety of sensor combinations.

The best performing sensors and the most efficient number of sensors were found using the same method as used for flowrate classification.

Table 8 shows that pressure classification yields poor results when only one or two sensors are used for the training and testing processes. The performance is best when four, five, or six sensors are used. Table 9, on the other hand, clearly shows that sensor 2 positively influences the classification, as it was the most abundant in the top performing 32% of classifiers.

Similar to the case of flow rate classifiers, sensor 2 may be the best choice because of its high gain, which further accentuates

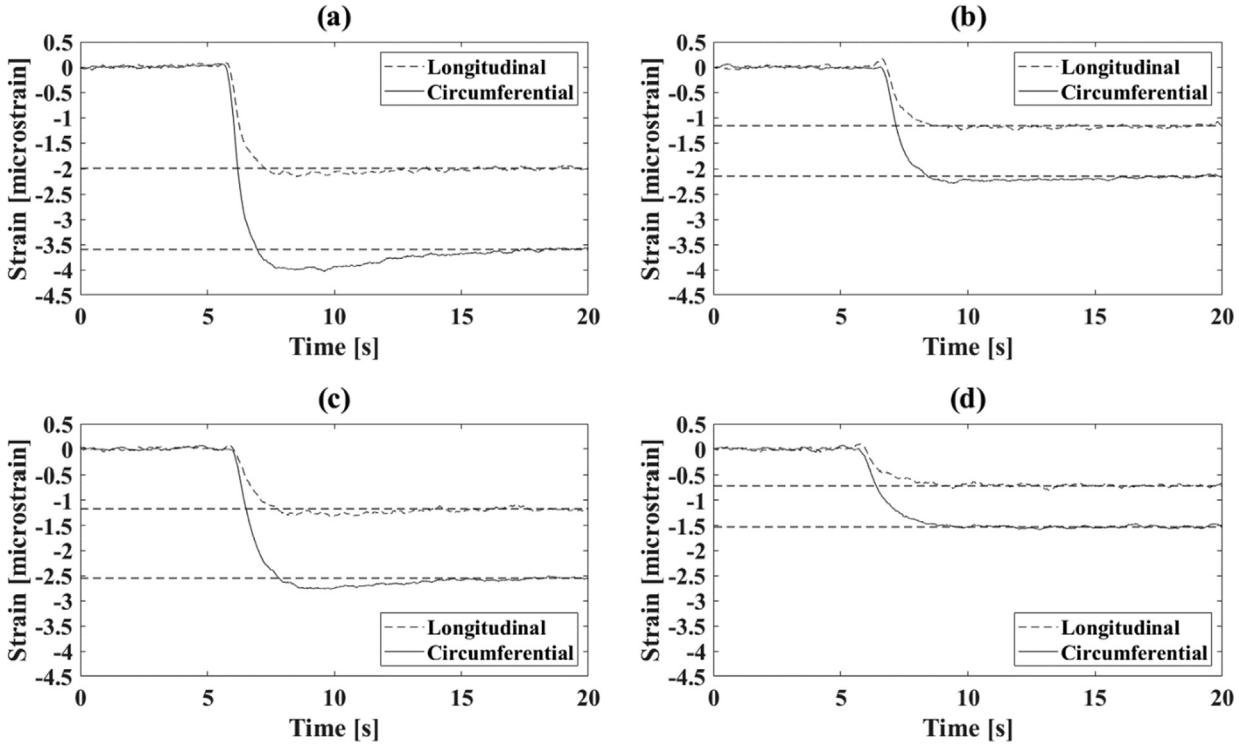

Fig. 15. Measured strain in the largest and smallest drop cases for a large leakage (a and b) and a small leakage (c and d).

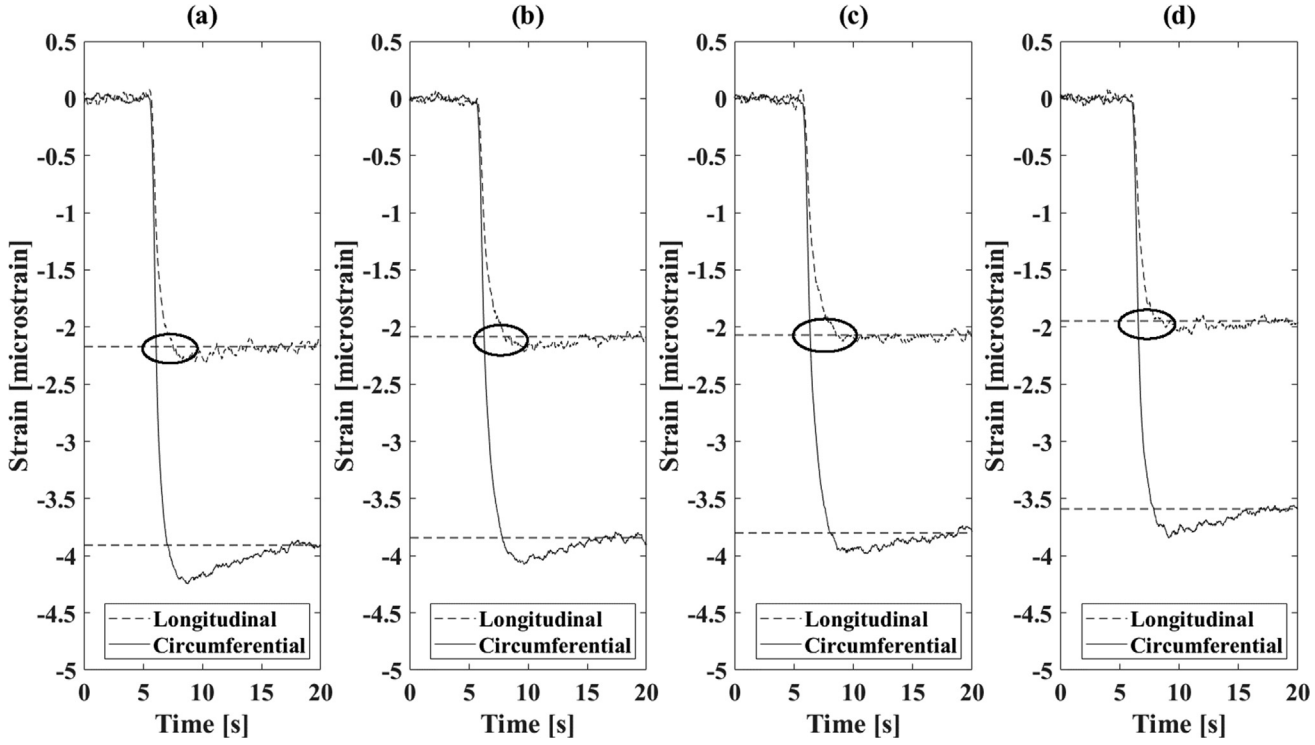

Fig. 16. Variations in leakage location at a pressure of 1 bar, a flow rate of 10 GPM, and a large leakage size for leakage locations (a) L1, (b) L2, (c) L3, and (d) L4.

the strain drop which was the main observed difference between cases with different flow rates.

5.5.3. Leakage size classification

Changes in leakage size had the most significant effects on the amplitude of the strain drop. However, in some extreme cases, changes in flow rate and pressure may be mistaken for a change in leakage size. Leakage size classifiers were therefore developed to overcome that uncertainty.

The performance of leakage size classifiers and the choice of the number of sensors and sensor configurations that work best for flow rate classification are presented in this section.

Table 3
Intercept times and time differences for sensors 1 and 2 recorded for different leakage scenarios.

Time intercepts for pressure variations at 15 GPM at leakage location L1 with a small leakage size			
	The intercept of sensor 1 in seconds (s)	The intercept of sensor 2 (s)	Time difference (s)
L1	7.525	6.104	1.421
L2	8.029	6.296	1.733
L3	8.552	6.374	2.178
L4	8.544	6.564	1.980
Time intercepts for pressure variations at 15 GPM at leakage location L4 with a small leakage size			
	The intercept of sensor 1 (s)	The intercept of sensor 2 (s)	Time difference (s)
L1	7.626	6.381	1.245
L2	7.547	6.212	1.335
L3	7.939	6.468	1.471
L4	8.152	6.552	1.600
Time intercepts for pressure variations at 15 GPM at leakage location L1 with a large leakage size			
	The intercept of sensor 1 (s)	The intercept of sensor 2 (s)	Time difference (s)
L1	8.147	6.775	1.372
L2	9.133	7.118	2.015
L3	9.118	6.775	2.343
L4	9.327	6.992	2.335
Time intercepts for pressure variations at 15 GPM at leakage location L4 with a large leakage size			
	The intercept of sensor 1 (s)	The intercept of sensor 2 (s)	Time difference (s)
L1	8.091	6.520	1.571
L2	9.284	6.675	2.609
L3	9.843	6.690	3.153
L4	9.081	6.497	2.584

Table 4
Ten best ranking results and result variation trends for flow rate classification.

Rank	Used sensors	Training accuracy (%)	Testing accuracy (%)
1	2, 3, 4, 6	95.26	88.33
2	2, 3, 6	94.93	87.77
3	2, 3, 6	95.53	87.77
4	2, 3, 4, 6	96.00	87.08
5	1, 2, 3, 4, 6	96.37	87.08
6	1, 2, 4	95.29	86.94
7	2, 6	94.39	86.94
8	1, 2, 3, 4, 6	95.81	86.80
9	1, 2, 3	96.49	86.80
10	1, 2, 3	95.13	86.80
20	2, 4	94.30	85.69
30	1, 2, 3, 4, 6	94.65	84.44
40	2, 4, 5	95.00	83.47
50	1, 3, 6	95.59	83.19
60	2, 4, 5	93.76	82.36
70	1, 3, 4	95.67	80.27
80	1, 3, 6	96.85	79.58
90	1, 3, 4, 5	95.46	79.02
100	1, 5	94.27	77.22
110	3, 4, 5	95.41	74.86
120	3	93.21	67.63
126	4	95.08	58.61

Table 10 shows that when using certain sensor combinations, the prediction accuracy can easily reach 100%. In the case of leakage size classification, accuracies ranged from 100% to 90%, with an average of 98.77%. These findings show that leakage size affects the signal in such a way that the classifier can easily detect it, regardless of the sensor combination used.

5.5.4. Leakage location classification

The pattern that emerged from changes in leakage location, as shown earlier, was unreliable for its sole use in the determination of leakage location. This is because of the higher number of exceptions found in these patterns. Leakage location classifiers were therefore developed for the ML algorithms to potentially identify other patterns that could help determine leakage locations.

The performance of leakage location classifiers and the choice of the number of sensors and sensor configurations that work best for flow rate classification are presented in this section.

Table 11 shows that when using certain sensor combinations, the prediction accuracy can reach 72%. The average testing accuracy for leakage location classification was 60.25% and ranged from 80.56 to 41.8%. These findings indicate that leakage location classification only works with certain sensor combinations. The best performing sensors and the most efficient number of sensors used were found using the same method as used for both pressure and flow rate classifications.

Table 12 shows that leakage location classification works better when a larger number of sensors is used. The only two instances (each using a different binary transformation) where all six sensors were used ended up in the top 32% of the results. Table 13, on the other hand, shows that sensor 4 was the most abundant sensor in the top performing 32% of the classifiers, closely followed by sensor 2. The sensitivity of sensor 2 may be a factor contributing to its good performance, whereas the effects of sensor 4 on yielding good results are possibly caused by its location on the testbed (closest to the middle).

Table 5
Number of appearances of instances with different numbers of sensors in the top-ranking flow rate classifiers.

Number of sensors	One sensor	Two sensors	Three sensors	Four sensors	Five sensors	Six sensors
Number of appearances	1	5	12	13	7	2
Number of appearances normalized	0.17	0.33	0.6	0.87	1.16	2

Table 6
Number of appearances of each sensor in the top-ranking flow rate classifiers.

Sensor	S1	S2	S3	S4	S5	S6
Number of appearances	20	40	24	23	15	24

Table 7
Ten best ranking results and result variation trends for pressure classification.

Rank	Used sensors	Training accuracy (%)	Testing accuracy (%)
1	1, 2, 3, 5	96.92	95.00
2	1, 2, 3, 5, 6	96.99	94.44
3	2, 3, 6	97.20	94.44
4	1, 2, 3	97.56	94.02
5	1, 2, 4, 6	96.52	94.02
6	2, 3, 4, 5, 6	96.85	94.02
7	1, 2, 3, 5, 6	95.90	93.88
8	1, 2, 3	96.90	93.75
9	1, 2, 6	95.85	93.75
10	2, 6	96.80	93.75
20	1, 2, 3, 4, 5, 6	97.04	93.33
30	2, 3, 5, 6	96.90	93.05
40	2, 5	94.80	92.22
50	1, 2, 3, 4, 5, 6	96.71	91.80
60	1, 5	96.14	91.25
70	1, 2, 4, 5, 6	96.47	90.83
80	1, 3, 4	96.11	89.58
90	3, 6	95.81	88.61
100	3, 4, 6	95.76	88.05
110	6	94.86	86.52
120	6	93.00	84.44
126	5	92.70	73.75

5.5.5. Classification of all parameters

It was mentioned earlier that the data were initially divided into two portions, one of which consisted of 10% of the data and was called global testing data. In this section, this portion was used to test the four classifiers. These tests were performed 10 times, and an analysis of the results and their repeatability is presented in this section.

Table 14 shows the 10 best ranking sensor configurations when it comes to all the classifiers. When each classifier was tested using global testing data, an average of their accuracies was calculated. This average was then averaged over the 10 iterations. The numbers presented in this table, therefore, reflect the quality of the sensor combinations in all the classification tasks.

The calculated average testing accuracies ranged between 88.06% and 68.9%, with an average of 82.87%. These findings indicate that some sensor combinations perform better than others. In this table, the combinations involving one or two sensors tended to score lower than the ones with more sensors. This result agrees with the observations above, where it was shown that these combinations were the least likely to be in the top 32% of the results for each classifier. Also, the standard deviation was low, with the highest being 3.04%, which indicates that the classifiers are generally consistent over 10 iterations.

6. Discussion

The recorded strain measurements present clear patterns that can be analyzed to determine pipe flow properties and leakage.

Table 9
Number of appearances of each sensor in the top-ranking pressure classifiers.

Sensor	S1	S2	S3	S4	S5	S6
Number of appearances	23	39	24	18	20	21

Table 10
Ten best ranking results and result variation trends for leakage size classification.

Rank	Used sensors	Training accuracy (%)	Testing accuracy (%)
1	1, 5	100	100
2	1, 5	100	100
3	2, 5	100	100
4	2, 5	100	100
5	3, 5	100	100
6	3, 5	100	100
7	5, 6	100	100
8	5, 6	100	100
9	1, 2, 5	100	100
10	1, 2, 5	100	100
20	2, 4, 5	100	100
30	1, 2, 3, 5	100	100
40	1, 4, 5, 6	100	100
50	1, 2, 3, 4, 5	100	100
60	1, 2, 3, 4, 5, 6	100	100
70	1, 2, 6	100	99.16
80	1, 2, 3, 6	100	98.75
90	1, 3, 4	99.07	98.33
100	4, 6	98.90	97.77
110	1, 6	98.54	97.36
120	3, 4	97.53	95.27
126	4	95.74	90.41

Table 11
Ten best ranking results and result variation trends for leakage location classification.

Rank	Used sensors	Training accuracy (%)	Testing accuracy (%)
1	2, 3, 4, 6	94.79	72.22
2	2, 4	96.59	70.69
3	1, 2, 4, 6	97.84	70.27
4	1, 2, 3, 4, 6	98.43	70.13
5	2, 3, 4, 5	98.09	69.44
6	2, 4, 5, 6	98.14	69.44
7	1, 2, 3, 4	95.48	69.16
8	1, 2, 4, 5, 6	98.50	69.02
9	1, 2, 3, 4, 5	98.43	68.88
10	1, 2, 4	94.65	68.47
20	1, 3, 4	95.19	67.08
30	1, 4	99.21	65.13
40	1, 2, 3, 5	98.52	63.47
50	1, 2, 3, 5, 6	99.06	62.63
60	2, 5, 6	97.88	60.69
70	1, 4, 5, 6	95.78	59.72
80	1, 2, 5	97.82	58.88
90	1, 3, 4, 5	95.29	58.05
100	1, 6	99.04	56.25
110	4	99.27	52.91
120	6	98.81	47.91
126	3	90.46	41.80

The patterns observed when any variable is changed remain stable, regardless of other flow and leakage properties. Strain drops are the most informative values, since three out of the four interrogated aspects directly influence the maximal strain drop. Each 5-GPM increment in flow rate yields a decrease of around 6% in max-

Table 8
Number of appearances of instances with different numbers of sensors in the top-ranking pressure classifiers.

Number of sensors	One sensor	Two sensors	Three sensors	Four sensors	Five sensors	Six sensors
Number of appearances	0	6	11	16	6	1
Number of appearances normalized	0	0.4	0.55	1.07	1	1

Table 12
Number of appearances of instances with different numbers of sensors in the top-ranking leakage location classifiers.

Number of sensors	One sensor	Two sensors	Three sensors	Four sensors	Five sensors	Six sensors
Number of appearances	1	5	14	13	5	2
Number of appearances normalized	0.17	0.33	0.7	0.86	0.83	2

Table 13
Number of appearances of each sensor in the top-ranking leakage location classifiers.

Sensor	S1	S2	S3	S4	S5	S6
Number of appearances	23	31	21	34	12	21

Table 14
Ten best ranking configurations for all the classifiers.

Rank	Used sensors	Average testing accuracy (%)	Standard deviation (%)
1	2, 3, 4, 6	88.06	1.54
2	1, 2, 3, 4, 6	87.92	1.77
3	1, 2, 4, 6	87.60	1.86
4	1, 2, 3, 4	87.57	2.40
5	2, 4, 6	87.43	2.08
6	1, 2, 3, 4, 6	86.94	1.77
7	2, 3, 4, 6	86.81	1.82
8	2, 4, 6	86.81	1.29
9	1, 2, 4	86.67	1.28
10	2, 3, 4, 5	86.60	2.23

imal strain drop. Pressure, however, affects maximal strain inversely, as each 0.2-bar increment results in an increase of strain drop by around 20%. The last property that affects maximal strain drop is the hole leakage size. The difference in maximal strain drop observed at instances with small and large leakages is an increase of around 55%, making this property the most prominent. Leakage location does not have any direct effect on maximal strain drop. However, leakage location affects the time it takes the sensors (based on their distance from the leakage) to reach steady-state.

Some of these properties are intertwined. For instance, the increase in leakage hole distance from the sensors dampens the effects of flow rate variations. The values presented here apply to the specific pipe used for testing. The behavior of the sensors towards any changes in flow properties was generally identical for longitudinal and circumferential sensors. However, the circumferentially mounted sensors had a higher gain and were hence more sensitive. This has helped reduce the effect of noise, which makes the collected signals reliable.

The classification of leakage size was easily achievable, with results reaching 100% accuracy with most sensor configurations. Pressure classification followed with good results, reaching 95% testing accuracy and a high accuracy for a large variety of sensor configurations. As for flow rate classification, the results were good, reaching 88% accuracy using the right settings. Finally, leakage location classification was performed with good results. Using the right settings, leakage location classification reached an accuracy of 72%. This task needs the use of a high number of sensors to achieve such results.

In the case of the classification of leakage location, there was a noticeable gap between the training and testing accuracies. The training accuracy, being much higher than the testing accuracy, indicated a case of classifier overfitting. However, training attempts that were performed with a smaller training data set (60% of the total data) still gave similar results. A possible explanation is that the pattern that emerged from the changes in the location of the leakage was too subtle and inconsistent to be used for classification. A longer testbed may help improve the accuracy and coherence of these numbers by accentuating and standardizing the differences between the signals generated by simulating leakages at different locations. Moreover, the pattern that emerged with a change in leakage location was related to time and, therefore, was weakened by the use of the moving mean filter that was necessary to reduce the noise in the data.

The results of pressure and flow rate classifications showed that it is possible to obtain better results using fewer sensors. In some cases in the literature, a greater number of sensors guaranteed better results, but the use of fewer sensors could provide a reliable and more cost-effective damage detection method [59]. However, in our specific case, a restrained number of sensors was capable of producing reliable results, as was the case with other instances in the literature [60–62], possibly because the pressure and flow rate trends were simple enough for classification using a restrained number of sensors. Increasing the inputs may only complicate the process. When the monitoring process is extended to a larger pipe, more sensors are needed, and sensor faults may be more frequent. Further, a sensor fault diagnostic method, like the one suggested by Huang et al. [63], can be applied to ensure the integrity of the sensor network.

By monitoring the temperature-compensated FBG strain data, it is possible to differentiate changes in the data caused by leakage from those caused by environmental factors and boundary conditions. This point was specifically studied by Chen et al. [42] who worked on developing SVM classifiers that were able to differentiate drops in pressure caused by a NPW from those caused by other external factors.

In terms of sensor performance, the tests showed that sensors placed in the circumferential direction enhanced the performance of all the developed classifiers. This finding is mainly due to their higher gain, which reduces the loss of valuable information in the data due to noise reduction and highlights the changes in the signals captured under different flow and leakage conditions. As for leakage localization, the position of the sensors played an important role in classification accuracy. The tests showed that using sensors placed near the center of the pipe tended to improve the accuracy of "leakage location" classifiers.

Broadening this monitoring method for real-life applications was facilitated with the use of FBG sensors, given their multiplexing capabilities that allow the daisy chaining of multiple sensors on each fiber strain. Further, FBG sensors are able to span long distances (up to 250 km) because of very low attenuation within the fiber [64]. Further, temperature-compensated FBG sensors can be used to negate the changes in the strain caused by the thermal expansion (environmental conditions) of the pipeline.

7. Conclusions

In this work, it is clear that pipeline leakages that occur under different operational conditions generate different amounts and variations in the strain patterns on the surface of the pipe. In all the tested scenarios, leakage always resulted in a drop of the strain measured on the surface of the pipe. The amplitude of that drop increased with an increase in pressure, a decrease in flow rate, and an increase in leakage size. The location of the leakage, however, had no effect on the amplitude of that drop but rather on the shape of the signals measured.

These observed patterns are sometimes subtle and can overlap. The shape and rate of the observed strain drop, although tied to pipe flow and leakage properties, will always have the same general aspect. These variations can rarely be differentiated by simple observations.

An SVM algorithm was used to implement multiclass classification to learn the patterns from the data collected. Using various sensor combinations, high accuracies were obtained, reaching 100% for leakage size classification, 95% for pressure classification, 88% for flow rate classification, and 72% for leakage location classification. Determining the size of the leakage resulted in the highest accuracy for the SVM classifiers, perhaps due to the fact that leakage size changes the signal most noticeably and consistently. SVM classifiers proved to be suitable for such applications, as they were able to determine the interrogated properties with good accuracy.

Based on the above results, FOS can be effectively used for the operational assessment of pipeline conditions. With the help of artificial intelligence, leakages can be localized, and a theoretical assessment of the damage can be performed from the moment the leak is detected.

CRediT authorship contribution statement

Michel Saade: Data curation, formal analysis, writing - original draft. **Samir Mustapha:** Conceptualization, formal analysis, funding acquisition, investigation, methodology, project administration, resources, supervision, writing-review and editing.

Declaration of Competing Interest

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

Acknowledgments

We acknowledge the financial support of the University Research Board at the American University of Beirut (Award #103604).

References

[1] G.A. Papadakis, Assessment of requirements on safety management systems in EU regulations for the control of major hazard pipelines, J. Hazard. Mater. 78 (1–3) (2000) 63–89, https://doi.org/10.1016/S0304-3894(00)00217-X.
[2] D. Lyons, Western European cross-country oil pipelines 30-year performance statistics, CONCAWE, 2002.
[3] J.L. Kennedy, Oil and Gas Pipeline Fundamentals, Pennwell Books, 1993.
[4] C. Sandberg, J. Holmes, K. McCoy, H. Koppitsch, The application of a continuous leak detection system to pipelines and associated equipment, IEEE Trans. Ind. Appl. 25 (5) (1989) 906–909, https://doi.org/10.1109/28.41257.
[5] H.A. Kishawy, H.A. Gabbar, Review of pipeline integrity management practices, Int. J. Press. Vessels Pip. 87 (7) (2010) 373–380, https://doi.org/10.1016/j.ijpvp.2010.04.003.
[6] A. Lubben, The Keystone oil spill no one's talking about will be nearly impossible to clean up, Vice, 2019.
[7] Bonterra has 70 workers cleaning up creek oil spill near Drayton Valley, 2019. www.cbc.ca/news/canada/calgary/bonterra-cleanup-oil-spill-alberta-washout-creek-drayton-1.5252378 (accessed 5-FEB 2020).
[8] J. Bohnen, Oil Spill Cleanup Underway in Garfield County, OK Energy Today (2019).
[9] T. Nye, Miller Grove oil spill releases over 8000 gallons of diesel into creek. News-Telegram, December 8 2019, 2019,
[10] J. Macpherson, A Keystone pipeline oil spill has affected nearly 10 times more land in North Dakota than first reported, Time (2019).
[11] P. Black, A review of pipeline leak detection technology, in: Pipeline Systems, Springer, 1992, pp 287–298
[12] A.D. Kersey, A review of recent developments in fiber optic sensor technology, Opt. Fiber Technol. 2 (3) (1996) 291–317, https://doi.org/10.1006/ofte.1996.0036.
[13] J.P. Lynch, C.R. Farrar, J.E. Michaels, Structural health monitoring: technological advances to practical implementations [scanning the issue], Proc IEEE 104 (8) (2016) 1508–1512, https://doi.org/10.1109/JPROC.2016.2588818.
[14] H. Jin, L. Zhang, W. Liang, Q. Ding, Integrated leakage detection and localization model for gas pipelines based on the acoustic wave method, J. Loss Prev. Process Ind. 27 (2014) 74–88, https://doi.org/10.1016/j.jlp.2013.11.006.
[15] J. Zhu, L. Ren, S.-C. Ho, Z. Jia, G. Song, Gas pipeline leakage detection based on PZT sensors, Smart Mater. Struct. 26 (2) (2017), https://doi.org/10.1088/1361-665X/26/2/025022 025022.
[16] R.A. Silva, C.M. Buiatti, S.L. Cruz, J.A.F.R. Pereira, Pressure wave behaviour and leak detection in pipelines, Comput. Chem. Eng. 20 (1996) S491–S496, https://doi.org/10.1016/0098-1354(96)00091-9.
[17] H. Jansen, M. Festen, Intelligent pigging developments for metal loss and crack detection, Insight-Wigston Then Northampton 37 (1995) 421.
[18] D. Mishra, K.K. Agrawal, A. Abbas, R. Srivastava, R. Yadav, PIG [Pipe Inspection Gauge]: An artificial dustman for cross country pipelines, Procedia Comput. Sci. 152 (2019) 333–340, https://doi.org/10.1016/j.procs.2019.05.009.
[19] F.B.I. Alnaimi, A.A. Mazraeh, K. Sahari, K. Weria, Y. Moslem, Design of a multi-diameter in-line cleaning and fault detection pipe pigging device, in: 2015 IEEE International Symposium on Robotics and Intelligent Sensors (IRIS), 2015. IEEE, pp 258–265. doi:10.1109/IRIS.2015.7451622.
[20] AbM Akib, Nb Saad, V. Asirvadam, Pressure point analysis for early detection system, in: Paper presented at the 2011 IEEE 7th International Colloquium on Signal Processing and its Applications, Penang, Malaysia, 4–6 March 2011, 2011.
[21] J.C. Martins, P. Seleghim, Assessment of the performance of acoustic and mass balance methods for leak detection in pipelines for transporting liquids, J. Fluids Eng. 132 (2010) 1, https://doi.org/10.1115/1.4000736.
[22] C.V.M. Ponce, G. Castañón, L. Cayuela, V. Puig, Model-based leak detection and location in water distribution networks considering an extended-horizon analysis of pressure sensitivities, J Hydroinformatics 16 (3) (2014) 649–670, https://doi.org/10.2166/hydro.2013.019.
[23] E. Hauge, O.M. Aamo, J.-M. Godhavn, Model based pipeline monitoring with leak detection, in: Seventh IFAC Symposium on Nonlinear Control Systems, vol. 1, 2007.
[24] T.-H. Yi, H.-N. Li, M. Gu, Wavelet based multi-step filtering method for bridge health monitoring using GPS and accelerometer, Smart Struct. Syst. 11 (4) (2013) 331–348, https://doi.org/10.12989/sss.2013.11.4.331.
[25] T.-H. Yi, H.-N. Li, H.-M. Sun, Multi-stage structural damage diagnosis method based on "energy-damage" theory, Smart Struct. Syst. 12 (3_4) (2013) 345–361, https://doi.org/10.12989/sss.2013.12.3_4.345.
[26] A. Mishra, Soni A Leakage detection using fibre optics distributed temperature sensing, in: 6th Pipeline Technology Conference, 2011.
[27] V. Spirin, M. Shlyagin, S. Miridonov, F.M. Jimenez, R.L. Gutierrez, Fiber Bragg grating sensor for petroleum hydrocarbon leak detection, Opt. Lasers Eng. 32 (5) (1999) 497–503, https://doi.org/10.1016/S0143-8166(00)00021-X.
[28] J.-R. Lee, H. Tsuda, Fiber optic liquid leak detection technique with an ultrasonic actuator and a fiber Bragg grating, Opt. Lett. 30 (24) (2005) 3293–3295, https://doi.org/10.1364/OL.30.003293.
[29] V. Nguyen, U. Dackermann, J. Li, M.M. Alamdari, S. Mustapha, P. Runcie, L. Ye, Damage identification of a concrete arch beam based on frequency response functions and artificial neural networks, Electron. J. Struct. Eng. 14 (1) (2015) 75–84.
[30] S. Mustapha, A. Braytee, L. Ye, Multisource data fusion for classification of surface cracks in steel pipes, J Nondestruct. Eval. Diagnost. Prognost. Eng. Syst. 1 (2) (2018), https://doi.org/10.1115/1.4038862.
[31] N.L.D. Khoa, B. Zhang, Y. Wang, W. Liu, F. Chen, S. Mustapha, P. Runcie, On damage identification in civil structures using tensor analysis, in: Pacific-Asia Conference on Knowledge Discovery and Data Mining, 2015, Springer, pp 459–471. doi:10.1007/978-3-319-18038-0-36.
[32] A. Khodayari-Rostamabad, J.P. Reilly, N.K. Nikolova, J.R. Hare, S. Pasha, Machine learning techniques for the analysis of magnetic flux leakage images in pipeline inspection, IEEE Trans. Magn. 45 (8) (2009) 3073–3084, https://doi.org/10.1109/TMAG.2009.2020160.
[33] X. Zhuang, S. Zhou, The prediction of self-healing capacity of bacteria-based concrete using machine learning approaches, Comput. Mater. Continua 59 (1) (2019), https://doi.org/10.32604/cmc.2019.04589.
[34] K.M. Hamdia, H. Ghasemi, X. Zhuang, N. Alajlan, T. Rabczuk, Computational machine learning representation for the flexoelectricity effect in truncated pyramid structures, Comput. Mater. Continua 59 (1) (2019), https://doi.org/10.15488/4769.
[35] M. Luo, K. Wang, Z. Cai, A. Liu, Y. Li, C.F. Cheang, Using imbalanced triangle synthetic data for machine learning anomaly detection, Comput. Mater. Continua 58 (1) (2019) 15–26, https://doi.org/10.32604/cmc.2019.03708.
[36] C.P. Huynh, S. Mustapha, P. Runcie, F. Porikli, Multi-class support vector machines for paint condition assessment on the Sydney Harbour Bridge using hyperspectral imaging, Struct. Monit. Maint. 2 (3) (2015) 181–197, https://doi.org/10.12989/smm.2015.2.3.181.
[37] H.Z. HosseinAbadi, R. Amirfattahi, B. Nazari, H.R. Mirdamadi, S.A. Atashipour, GUW-based structural damage detection using WPT statistical features and multiclass SVM, Appl. Acoust. 86 (2014) 59–70, https://doi.org/10.1016/j.apacoust.2014.05.002.
[38] S.-F. Yuan, F.-L. Chu, Support vector machines-based fault diagnosis for turbo-pump rotor, Mech. Syst. Signal Pr. 20 (4) (2006) 939–952, https://doi.org/10.1016/j.ymssp.2005.09.006.
[39] A. Widodo, E.Y. Kim, J.-D. Son, B.-S. Yang, A.C. Tan, D.-S. Gu, B.-K. Choi, J. Mathew, Fault diagnosis of low speed bearing based on relevance vector machine and support vector machine, Expert Syst. Appl. 36 (3) (2009) 7252–7261.

[40] T. Oh, J. Kim, A. Zhang, C. Lee, S. Park, Concrete strength evaluation in an early-age curing process using SVM with ultrasonic harmonic waves, Insight-Non-Destruct. Test. Cond. Monit. 58 (11) (2016) 609–616, https://doi.org/10.1784/insi.2016.58.11.609.
[41] S. Park, D.J. Inman, J.-J. Lee, C.-B. Yun, Piezoelectric sensor-based health monitoring of railroad tracks using a two-step support vector machine classifier, J. Infrastruct. Syst. 14 (1) (2008) 80–88, https://doi.org/10.1061/ASCE1076-0342200814:180.
[42] C. Huali, Y. Hao, C. Lv, S. Hongyu, Application of support vector machine learning to leak detection and location in pipelines, in: Proceedings of the 21st IEEE Instrumentation and Measurement Technology Conference (IEEE Cat. No. 04CH37510), 2004. IEEE, pp. 2273–2277. doi:10.1109/IMTC.2004.1351546
[43] Y. Ying, J. Harley, J. Garrett, H. James, Y. Jin, I.J. Oppenheim, J. Shi, L. Soibelman, Applications of machine learning in pipeline monitoring, in: Paper presented at the International Workshop on Computing in Civil Engineering 2011, 2012,
[44] Z. Qu, H. Feng, Z. Zeng, J. Zhuge, S. Jin, A SVM-based pipeline leakage detection and pre-warning system, Measurement 43 (4) (2010) 513–519, https://doi.org/10.1016/j.measurement.2009.12.022.
[45] W.-Y. Chuang, Y.-L. Tsai, L.-H. Wang, Leak detection in water distribution pipes based on CNN with mel frequency cepstral coefficients, in: Proceedings of the 2019 3rd International Conference on Innovation in Artificial Intelligence, 2019. ACM, pp 83–86. doi:10.1145/3319921.3319926.
[46] R.B. Santos, M. Rupp, S.J. Bonzi, A.M.F. Fileti, Comparison between multilayer feedforward neural networks and a radial basis function network to detect and locate leaks in pipelines transporting gas, Chem. Eng. Trans. 32 (1375) (2013), https://doi.org/10.3303/CET1332230 e1380.
[47] S. Sfarra, C. Ibarra-Castanedo, F. Lambiase, D. Paoletti, A. Di Ilio, X. Maldague, From the experimental simulation to integrated non-destructive analysis by means of optical and infrared techniques: results compared, Meas. Sci. Technol. 23 (11) (2012), https://doi.org/10.1088/0957-0233/23/11/115601 115601.
[48] Y.-J. Rao, In-fibre Bragg grating sensors, Meas. Sci. Technol. 8 (4) (1997) 355, https://doi.org/10.1088/0957-0233/8/4/002.
[49] X. Ye, T. Jin, C. Yun, A review on deep learning-based structural health monitoring of civil infrastructures, Smart Struct. Syst. 24 (5) (2019) 567–585, https://doi.org/10.12989/sss.2019.24.5.567.
[50] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (3) (1995) 273–297.
[51] C.-W. Hsu, C.-J. Lin, A comparison of methods for multiclass support vector machines, ITNN 13 (2) (2002) 415–425.
[52] C.R. Farrar, K. Worden, Supervised learning – classification and regression, in: Structural Health Monitoring. John Wiley & Sons, Ltd, 2012, pp. 361–401. doi:10.1002/9781118443118.ch11.
[53] H. Byun, S.-W. Lee, A survey on pattern recognition applications of support vector machines, Int. J. Pattern Recognit Artif Intell. 17 (03) (2003) 459–486, https://doi.org/10.1142/S0218001403002460.
[54] T.H. Yi, H.N. Li, M. Gu, Optimal sensor placement for structural health monitoring based on multiple optimization strategies, Struct. Des. Tall Special Build. 20 (7) (2011) 881–900, https://doi.org/10.1002/tal.712.
[55] T.-H. Yi, H.-N. Li, M. Gu, A new method for optimal selection of sensor location on a high-rise building using simplified finite element model, Struct. Eng. Mech. 37 (6) (2011) 671–684, https://doi.org/10.12989/sem.2011.37.6.671.
[56] T. Jiang, L. Ren, Z. Jia, D. Li, H. Li, Application of FBG based sensor in pipeline safety monitoring, Appl. Sci. 7 (6) (2017) 540.
[57] Z. Jia, L. Ren, H. Li, W. Sun, Pipeline leak localization based on FBG hoop strain sensors combined with BP neural network, Appl. Sci. 8 (2) (2018) 146.
[58] Natural Resources Canada, 2020. https://www.nrcan.gc.ca/home (accessed 4-MAY 2020).
[59] S.-W. Lin, T.-H. Yi, H.-N. Li, L. Ren, Damage detection in the cable structures of a bridge using the virtual distortion method, J. Bridge Eng. 22 (8) (2017) 04017039, https://doi.org/10.1061/(ASCE)BE.1943-5592.0001072.
[60] I. Ahmad, M. Hussain, A. Alghamdi, A. Alelaiwi, Enhancing SVM performance in intrusion detection using optimal feature subset selection based on genetic principal components, Neural Comput. Appl. 24 (7–8) (2014) 1671–1682, https://doi.org/10.1007/s00521-013-1370-6.
[61] D.S. Kim, H.-N. Nguyen, J.S. Park, Genetic algorithm to improve SVM based network intrusion detection system, in: 19th International Conference on Advanced Information Networking and Applications (AINA'05), vol. 1 (AINA papers). IEEE, 2005, pp 155–158. doi:10.1109/AINA.2005.191.
[62] Z. Jia, L. Ren, H. Li, T. Jiang, W. Wu, Pipeline leakage identification and localization based on the fiber Bragg grating hoop strain measurements and particle swarm optimization and support vector machine, Struct. Contr. Health Monit. 26 (2) (2019), https://doi.org/10.1002/stc.2290 e2290.
[63] H.-B. Huang, T.-H. Yi, H.-N. Li, Bayesian combination of weighted principal-component analysis for diagnosing sensor faults in structural monitoring systems, J. Eng. Mech. 143 (9) (2017) 04017088, https://doi.org/10.1061/(ASCE)EM.1943-7889.0001309.
[64] M. Fernandez-Vallejo, S. Rota-Rodrigo, M. Lopez-Amo, Remote (250 km) fiber Bragg grating multiplexing system, Sensors 11 (9) (2011) 8711–8720, https://doi.org/10.3390/s110908711.