



Data-Driven Inline Leak Detection for Pipelines Using Flow-Induced Acoustics Analysis

Saravanabalaji M¹, Shakthi Raagavi S², Yogesh K³, Sneha S⁴, Hariharasudhan P⁵

¹ Assistant Professor, Electronics and Instrumentation Engineering, Kumaraguru College of Technology, Coimbatore, Tamilnadu, India.

^{2,3,4,5} Electronics and Instrumentation Engineering, Kumaraguru College of Technology, Coimbatore, Tamilnadu India.

How to cite this paper:

Saravanabalaji M¹, Shakthi Raagavi S²,
Yogesh K³, Sneha S⁴, Hariharasudhan P⁵,
"Data-Driven Inline Leak Detection for
Pipelines Using Flow-Induced Acoustics
Analysis", IJIRE-V5I02-200-207.

Copyright © 2024 by author(s) and
5th Dimension Research Publication.

This work is licensed under the Creative
Commons Attribution International License
(CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>

Abstract: Fluid and water distribution networks are essential to the modern world. However, these systems are prone to leaks, which can lead to significant water loss, damage to infrastructure, and environmental pollution. The proposed solution makes use of Acoustic Emission sensors placed in discrete locations in the pipeline which measures the sound in the pipeline caused by the flow of fluids. Computation models are used to deduce the location from the input provided by the sensors. In case of leak, the leak is localized through cross correlation and TDOA methods. This solution is particularly developed for water distribution pipelines.

Keyword: Acoustic data analysis, Data-driven models, Cross-Correlation, Time Difference of Arrival (TDOA)

1.INTRODUCTION

There are many traditional leak detection methods for water pipelines. The Standing Wave Difference Method (SWDM) has been identified as a promising technique for leak detection in water transmission systems [1]. Additionally, the analysis of leakage acoustic signals is effective in detecting small leaks in water supply pipelines [2]. Furthermore, the use of wireless sensor networks and noisy pressure sensor data has been explored for locating leaks in water pipelines, highlighting the importance of real-time monitoring and data-driven approaches [3]. However, traditional leak detection methods face challenges such as the requirement for expensive measurements of total pipeline flow and other variables at multiple points, which may lead to unsatisfactory results due to the neglect of transient changes in the product variables [4]. Moreover, the complex vibration signals in gas pipelines pose challenges for traditional leak detection methods, as valuable leakage information may be concealed within the data [5]. In recent years, data-driven approaches have gained attention for leak detection in pipeline systems. These approaches include the use of artificial neural networks, recurrent neural networks, and support vector machines for identifying and localizing leaks in pipelines [6]. Additionally, the integration of advanced techniques such as particle swarm optimization, principal component analysis, and convolutional neural networks has shown promise in pipeline leak detection [7] [8]. Furthermore, the development of novel sensor technologies, such as fiber optic distributed sensors, has enabled real-time monitoring and detection of pipeline leaks, demonstrating the potential for continuous and accurate leak detection [9]. Additionally, the utilization of dynamic pressure waves and transient flow modeling has been proposed for leak detection and location in gas pipelines, highlighting the importance of innovative approaches for different types of pipelines [10][11]. The development of an integration method using kernel principal component analysis and cascade support vector data description for pipeline leak detection emphasizes the importance of advanced data analytics and feature extraction methods in the detection system [12]. In conclusion, traditional leak detection methods have limitations in effectively identifying and localizing leaks in water and gas pipelines. The integration of data-driven approaches, advanced sensor technologies, and innovative modeling techniques has shown promise in addressing these limitations and improving the accuracy and efficiency of leak detection in pipeline systems. Traditional leak detection methods have limitations in effectively identifying and localizing leaks in water and gas pipelines. The integration of data-driven approaches, advanced sensor technologies, and innovative modeling techniques has shown promise in addressing these limitations and improving the accuracy and efficiency of leak detection in pipeline systems. Data-driven leak detection system in water pipelines using advanced sensor technologies, data-driven models, real-time monitoring capabilities, advanced signal processing techniques, and robust data validation methods are essential for ensuring the accuracy, reliability, and efficiency of leak detection in water distribution networks.

II. MATERIAL AND METHODS

Specifications and Initial Study

A. Pipeline setup

A steel pipeline, along with a water tank and pump system, has been installed in the laboratory.



Fig. 1. Pipeline setup

The pipeline setup in Fig. 1 has artificially induced leaks and hydrophone sensors mounted in it

B. Simulation in Comsol

A 3D model of the pipeline setup has been created and simulated using COMSOL using Turbulent Flow module with constant inlet velocity of 1m/s to analyze the variations in pressure within the pipe flow under leak conditions. Simulation of the fluid flow in pipeline is done in comsol with the following parameters:

- I. Time-dependent, Stationary Study
- II. Turbulent Flow – Reynold's number >2000
- III. Mesh – Extremely coarse (To reduce the computation time)
- IV. Boundary conditions:
Inlet – Velocity (1 m/s)
Outlet – Pressure (Atmospheric Pressure)

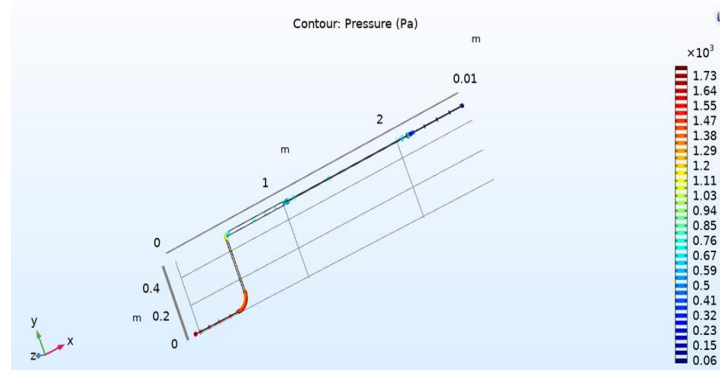


Fig. 2. Pressure Plot in Comsol

It is observed from the pressure plot (Fig. 2) that pressure abruptly decreases at section 2 where the leak is introduced.

C. Data Acquisition from Accelerometer sensor

Vibration data under both leak and non-leak conditions is gathered utilizing a Dytran Accelerometer sensor. This sensor is interfaced with a NI 9234 Data Acquisition (DAQ) module, specifically designed for sound and vibration measurements, operating within the LabView environment. The sensor is attached on the surface of the pipeline using magnet.



Fig. 3. Accelerometer sensor

Fig. 3 shows the accelerometer sensor fitted on the surface of the pipeline.

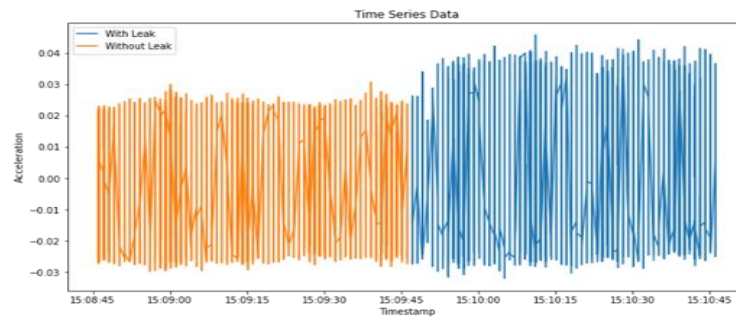


Fig. 4. Plot of Vibration data

From Fig. 4., it is observed that there is a clear difference in the amplitude of the vibration data for leak and non-leak conditions. However, the collected data is significantly susceptible to interference from external noises. Therefore, opting for a Hydrophone sensor is better, as it is specifically designed for recording underwater sounds, thus minimizing the impact of external disturbances.

D. Data Acquisition from Hydrophone sensor

Acoustic data under both leak and non-leak conditions is gathered utilizing two Hydrophone h1c sensors. It is an IEPE sensor that works based on piezoelectric principle. This sensor is interfaced with a NI 9234 Data Acquisition (DAQ) module, specifically designed for sound and vibration measurements, operating within the LabView environment. The sensors are inserted inside the pipeline using a T-joint. Two hydrophone sensors are fitted on either side of the leak.



Fig. 5. Hydrophone sensor

Fig. 5 shows the hydrophone sensor fitted inside a T-shaped joint in the pipeline



Fig 6. Leak

Fig. 6 shows artificially induced leak using a control valve

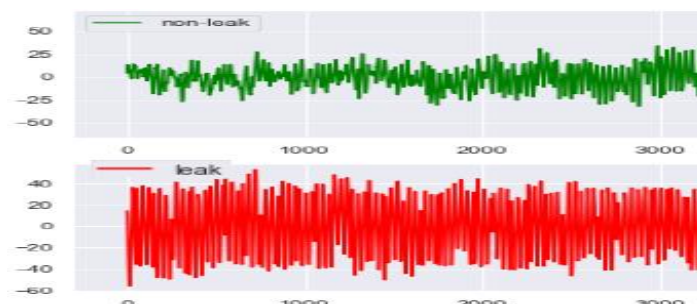


Fig. 7. Plot of Acoustic data

Fig. 7 represents the time-domain acoustic signals for leak and non-leak conditions.

Proposed Architecture

A. Methodology

1. Model Training

- Collection of Acoustic data from hydrophone sensor for leak and non-leak conditions
- Conversion of the data to frequency domain using the Fast Fourier Transform (FFT)
- Train the data using ensemble modeling
- Cross-validate the trained model ensemble

2. Detection and Localization

- Collection of acoustic data from the sensor
- Conversion of Time Domain Signal to Frequency Domain Signal using FFT
- Dimension of the Transformed input data is reduced using PCA
- This data is tested for leaks using the trained model
- In case of leak, the leak is located using TDOA and Cross-Correlation, else the cycle is continued.

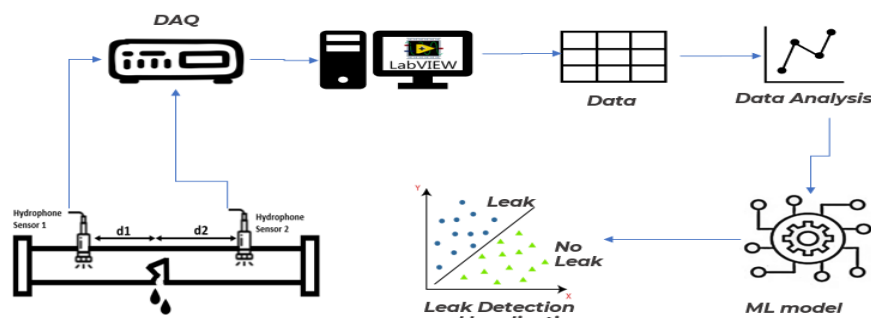


Fig. 8. Outline of the proposed solution

B. Fourier Transform

Time domain Acoustic signal collected from the hydrophone sensors is converted into frequency domain signal using Fast Fourier Transform (FFT).

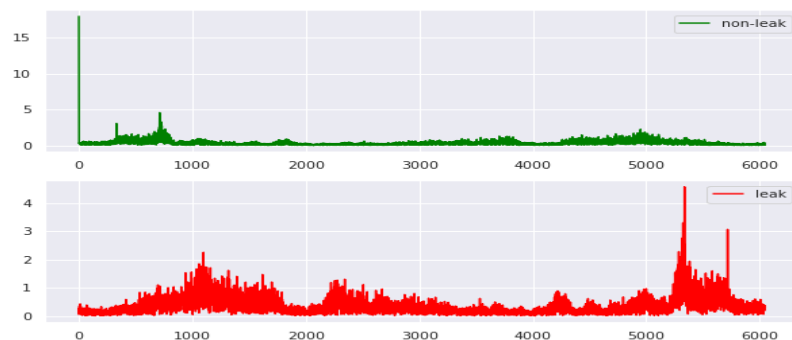


Fig. 9. Frequency domain acoustic signal

Fig. 9 represents Fast Fourier Transformed acoustic signal (Frequency domain acoustic signal).

The following four features are extracted from the frequency domain signal for analysis.

Features	Non-Leak	Leak
Spectral Bandwidth	12803.49	13863.15
Spectral Centroid	-0.01337	-0.0051
RMS Energy	8.2174	9.1650

Table 1 – Features for leak and non-leak

Table 1 suggests that in the presence of a leak, there is a broader spread or distribution of energy across different frequency components compared to the non-leak condition.

A higher value of spectral centroid indicates that the average frequency of the signal is shifted towards higher frequencies in the presence of a leak.

The higher RMS energy indicates an increase in the overall energy content or intensity in the presence of a leak. This

suggests that the signal is more energetic during leak conditions, possibly due to the presence of additional acoustic activity associated with the leak.

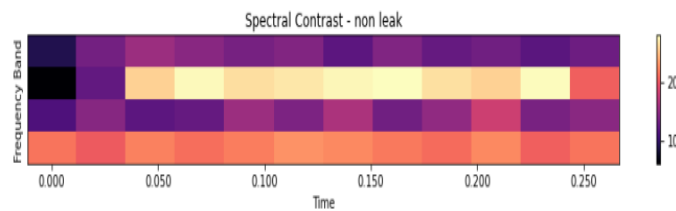


Fig. 10. Spectral Contrast for Non-Leak data

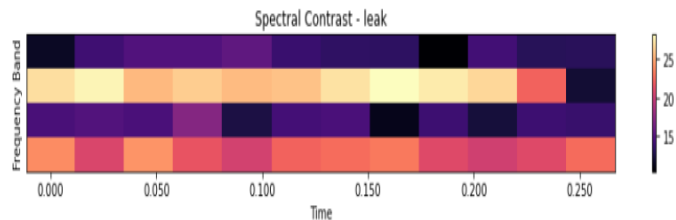


Fig. 11. Spectral Contrast for Leak data

C. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a dimensionality reduction technique used to transform high-dimensional data into lower-dimension while preserving most of the original variance in the data.

The original data shape is (84, 6050). This projection reduces the dimensionality of the data from features to $n_components$, where features is the original number of features in the dataset.

After PCA, the first 17 principal components capture a significant portion of the variance in the original data, hence the dimensionality is reduced to 17.

```
n_components: 1, variance: 0.7855233316738508
n_components: 2, variance: 0.9660568781772643
n_components: 3, variance: 0.9776484461591877
n_components: 4, variance: 0.9811624426252017
n_components: 5, variance: 0.9828171897505092
n_components: 6, variance: 0.9842063560319365
n_components: 7, variance: 0.9851553113127782
n_components: 8, variance: 0.9859436033281479
n_components: 9, variance: 0.986602360900876
n_components: 10, variance: 0.9872057288726704
n_components: 11, variance: 0.9877366215611472
n_components: 12, variance: 0.9882459338723294
n_components: 13, variance: 0.9887184420889145
n_components: 14, variance: 0.9891423349475646
n_components: 15, variance: 0.9895631402733728
n_components: 16, variance: 0.9899734022351842
n_components: 17, variance: 0.9902831697871904
```

Fig. 12. PCA Data

Fig. 12 shows the first 17 principal components capturing a maximum portion of variance in the data (0.99).

D. Model Training

The final PCA data is trained using ensemble modeling. Ensemble modeling is a machine learning technique that uses multiple models to predict an outcome. Ensemble methods have higher predictive accuracy than individual models. The model is trained according to the following steps.

1. The data is split into 10 folds for cross-validation using KFold.
2. Iterate over each fold:
 - i. The data is split into training and testing sets.
 - ii. The following classification models (Logistic Regression, Support Vector Machine with RBF kernel, AdaBoost, Random Forest) are trained on the training dataset
 - iii. Predictions are made on the test data using each model.
 - iv. Raw predicted probabilities for each class are calculated.
 - v. Predictions from all models are combined using weighted averaging.
 - vi. The combined raw probabilities are converted to binary predictions using a threshold of 0.6.
 - vii. The performance metrics (accuracy, precision, recall, F1 score) of the combined predictions are evaluated.
3. The performance metrics of different combination of classification models with different weights are evaluated. Of which the following combination of classification models worked with better accuracy.


```

preds_raw = (0.7*logit_preds_raw + 0.1*svc_linear_preds_raw + 0.1*ad_preds_raw +
             0.1*rf_preds_raw) / 1
preds_labels = (preds_raw > 0.6).astype(int)

```

Fig. 13. Combination of ensemble models

	fold	accuracy	precision	recall	f1_score
count	10.00000	10.000000	10.000000	10.000000	10.000000
mean	5.50000	0.986667	0.988194	0.986607	0.987189
std	3.02765	0.023307	0.024942	0.028313	0.022244
min	1.00000	0.933333	0.937500	0.928571	0.937500
25%	3.25000	0.975000	1.000000	1.000000	0.978571
50%	5.50000	1.000000	1.000000	1.000000	1.000000
75%	7.75000	1.000000	1.000000	1.000000	1.000000
max	10.00000	1.000000	1.000000	1.000000	1.000000

Fig. 14. Performance metrics of the ensemble model

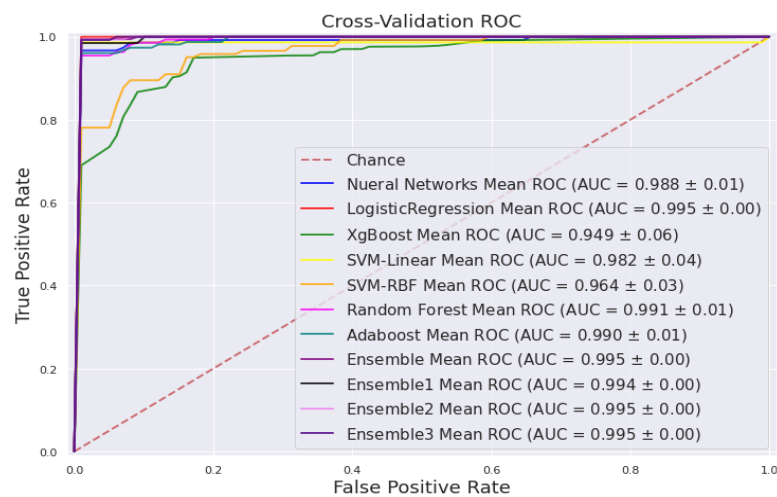


Fig. 15. ROC curve for the ensemble model

From Fig. 15., it is observed that the ROC (Receiver Operating Characteristic) curve of the ensemble mean is closer to the top-left corner of the plot and the values of AUC (Area Under the Curve) of ensemble mean being closer to 1(0.99) which are the criteria for a better performing model.

E. Leak Localization

Leak localization is done using cross-correlation and Time Difference of Arrival (TDOA) methods. Two hydrophones are set on the pipeline on either side of the artificially induced leak as below.

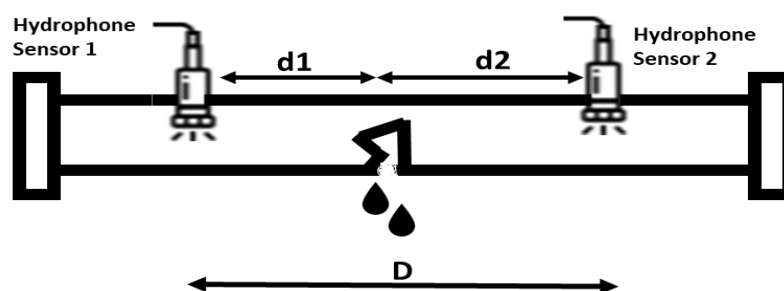


Fig. 16. Sensor position on the pipe with leak

In Fig. 16.

d1 – distance between sensor 1 and the leak

d2 – distance between sensor 2 and the leak

D – Distance between the two sensors

The signals from the leak arrive at the sensors at different times due to the varying distances between the sensors and the leak location, as well as the speed of signal propagation through the medium. Time difference (lag) between sensor 1 and sensor 2 signals is found using cross-correlation.

Cross-correlation measures the similarity between two signals as a function of the lag of one relative to the other. The cross-correlation function is computed by sliding one signal (the template or reference signal) across the other signal (the input or target signal) and computing the dot product at each lag. The lag where the cross-correlation is maximum corresponds to the time delay between the two signals.

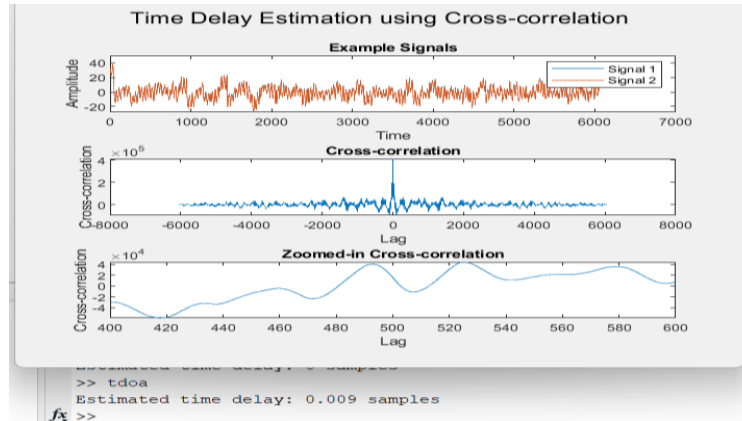


Fig. 17. Cross correlation Graph

The time delay is obtained in samples because the cross-correlation function is often computed discretely over time, with each sample representing a discrete time step. Therefore, the lag at which the correlation is maximum indicates the number of samples by which one signal is delayed relative to the other. Time delay in samples is converted into time delay in seconds using the following formula,

$$\text{Time delay (seconds)} = \frac{\text{Time delay (samples)}}{\text{Sampling frequency}}$$

Sound propagation speed in the pipeline (c) is calculated using the following formula,

$$C = C_f \left(1 + \frac{2Ba}{Eh} \right)^{-1/2}$$

Where,

B – Fluid Bulk Modulus of Elasticity (in GPa)

A – Pipe radius (in metre)

h – Pipe wall thickness (in metre)

E – Young's modulus of the pipe wall material (in GPa)

c_f – Speed of sound in air (in m/s)

The distance of the leak from sensor 1 (d1) is calculated using the following formula,

$$d_1 = \frac{D + cT_0}{2}$$

Where,

T_0 – Time delay in seconds

c – Sound propagation speed in the given medium

D – Distance between the two sensors

III.RESULT

In this study, effect in the water flow due to leak is studied by simulating the 3D model of the pipeline in Comsol using CFD (Computational Fluid Dynamics) and FEA (Finite Element Analysis). Variation in the real-time signal is tested by studying the obtaining frequency domain features. Acoustic data obtained from the hydrophone sensor is trained using ensemble learning and tested which gives an average accuracy of 0.9. Position of the leak is localized by Cross Correlation and TDOA methods which estimates the distance of the leak from Sensor 1. The project successfully developed a machine learning-based approach for leak localization using acoustic data, dimensionality reduction, model training, and ensemble techniques to

achieve high predictive accuracy and reliability. From Fig. 15, it is inferred that the ensemble model achieved superior performance compared to individual models, as evidenced by higher accuracy, precision, recall, F1-score, and AUC. The final model demonstrated robust leak localization capability, providing accurate predictions with low false positive and false negative rates.

IV. DISCUSSION

Simulation of water flow in a 3D pipeline in Comsol using Computational Fluid Dynamics and Finite Element Analysis helped in understanding the variation of parameters w.r.t distance from the leak. It played a major role in choosing the optimal sensor location. Vibration data from Dytran Accelerometer and acoustic data from hydrophone are tested for leak and non-leak conditions and for noise filtering of which hydrophone proved to be more effective. The acoustic data is converted to frequency domain for stability and trained with different ML algorithms and ensemble learning proved to be effective in better accuracy and performance of the model. The acoustic data from the sensors on either side of the leak helped in leak localization accurately using cross correlation and TDOA methods. Though the proposed methodology proved to be effective in pipe leakage detection and localization, this methodology can further be made promising by integrating the following features - developing methods for reducing the impact of background noise on acoustic leak detection, collecting more real-world data to evaluate the performance of the proposed leak detection system, predicting leaks even before they occur by analyzing the stress on the pipe material using FEA (Finite Element Analysis) and by computing the severity of the leak.

V. CONCLUSION

Our methodology offers a reliable and efficient solution for leak localization in water pipelines, contributing to improved pipeline maintenance, reduced unnecessary water loss and risk mitigation strategies. The proposed methodology enhances water conservation efforts by swiftly identifying and addressing leaks, minimizing water wastage and associated costs. It facilitates early detection of leaks, mitigating potential damage to infrastructure and properties, thus ensuring operational efficiency and safety.

References

1. Covas, D., Ramos, H., & Almeida, A. (2005). Standing wave difference method for leak detection in pipeline systems. *Journal of Hydraulic Engineering*, 131(12), 1106-1116.
2. Li, Y., Zhou, Y., Fu, M., Zhang, F., Chi, Z., & Wang, W. (2021). Analysis of propagation and distribution characteristics of leakage acoustic waves in water supply pipelines. *Sensors*, 21(16), 5450.
3. Mysorewala, M., Sabih, M., Cheded, L., Nasir, M., & Ismail, M. (2015). A novel energy-aware approach for locating leaks in water pipeline using a wireless
4. Fouad, M., Al-AWADI, H., Moustafa, F., & Nawara, M. (2008). Leak detection and localization of the pipeline systems. *The International Conference on Applied Mechanics and Mechanical Engineering*, 13(13), 431-448.
5. Sun, J., Liu, Z., Wen, J., & Qiao, Y. (2022). Leakage aperture identification of natural gas pipeline based on compressed acquisition and dsae. *Structural Health Monitoring*, 22(4), 2856-2867.
6. Jia, Z., Ren, L., Li, H., Jiang, T., & Wu, W. (2018). Pipeline leakage identification and localization based on the fiber bragg grating hoop strain measurements and particle swarm optimization and support vector machine. *Structural Control and Health Monitoring*, 26(2), e2290.
7. Bohorquez, J., Lambert, M., Alexander, B., Simpson, A., & Abbott, D. (2022). Stochastic resonance enhancement for leak detection in pipelines using fluid transients and convolutional neural networks. *Journal of Water Resources Planning and Management*, 148(3).
8. Yan, W., Liu, W., Bi, H., Jiang, C., Yang, D., Sun, S. ... & Sun, Y. (2023). Acoustic detection and localization of gas pipeline leak based on residual connection and one-dimensional separable convolutional neural network. *Transactions of the Institute of Measurement and Control*, 45(14), 2637-2647.
9. Hai-ou, S., Zhu, Y., & Lang, X. (2022). A small leak detection and localization method for oil pipelines based on improved robust principle component analysis. *Journal of Physics Conference Series*, 2264(1), 012021.
10. Yan, W., Liu, W., Bi, H., Jiang, C., Yang, D., Sun, S., & Sun, Y. (2023). Acoustic detection and localization of gas pipeline leak based on residual connection and one-dimensional separable convolutional neural network. *Transactions of the Institute of Measurement and Control*, 45(14), 2637-2647.
11. Civan, F. and Balda, K. (2013). Application of mass balance and transient flow modeling for leak detection in liquid pipelines.
12. Zhou, M., Zhang, Q., Liu, Y., Sun, X., Cai, Y., & Pan, H. (2019). An integration method using kernel principal component analysis and cascade support vector data description for pipeline leak detection with multiple operating modes.