Systems & Control Transactions

Research Article - Peer Reviewed Conference Proceeding
ESCAPE 35 - European Symposium on Computer Aided Process Engineering
Ghent, Belgium. 6-9 July 2025
Jan F.M. Van Impe, Grégoire Léonard, Satyajeet S. Bhonsale,
Monika E. Polańska, Filip Logist (Eds.)

PSE PRESS

Developing a Digital Twin System Based on a Physics-informed Neural Network for Pipeline Leakage Detection

Wei-Shiang Lin[a], Yi-Hsiang Cheng[b], Zhen-Yu Hung[b], and Yuan Yao[a]*

[a] Department of Chemical Engineering, National Tsing Hua University, Hsinchu, Taiwan
[b] Material and Chemical Research Laboratories, Industrial Technology Research Institute, Hsinchu, Taiwan
* Corresponding Author: yyao@mx.nthu.edu.tw.

ABSTRACT

As the demand for resources continues to grow, pipelines have become critical for transporting water, fossil fuels, and chemicals. Monitoring pipeline systems is essential, as leaks can lead to severe environmental damage and safety hazards. This study aims to develop a pipeline leakage detection system based on digital twin technology and Physics-Informed Neural Networks (PINNs). By embedding physical principles, such as the continuity and momentum equations derived from the Navier-Stokes equation, into the neural network's loss function, the model can predict pressure and flow dynamics with high accuracy while adhering to physical constraints. PINNs are particularly advantageous as they require minimal data, maintain physical consistency, and provide reliable interpretations, making them well-suited for addressing pipeline safety challenges. The model is designed to simulate fluid dynamics under normal operating conditions, with deviations in prediction errors signaling potential leaks. Statistical analysis of these errors is used to define control limits, establish rejection regions, and create control charts for leak detection. The detection system is further validated using field data to ensure reliability. By combining physical modeling and neural networks, this approach enhances the accuracy and interpretability of leakage detection, laying a solid foundation for an efficient pipeline monitoring solution.

Keywords: Pipeline leakage detection, Physics-informed neural networks, Surrogate model, Industrial safety

INTRODUCTION

As modern society's demand for energy and resources continues to grow, pipeline transportation has become a critical foundation for industrial and domestic supply. However, pipeline leaks not only pose a significant threat to the environment through severe pollution but also endanger the safety of nearby residents and industrial surroundings. Furthermore, such incidents can compromise the quality of transported materials, with ruptured pipelines potentially leading to product contamination. Therefore, the monitoring and control of pipeline systems are of utmost importance.

Current pipeline leakage detection technologies can be broadly classified into two categories: hardware-based methods and software-based methods [1]. Hardware-based methods utilize equipment to detect leaks and their locations. Common techniques include acoustic sensing, fiber optic cables, ground-penetrating radar, gas injection, and infrared thermography. For example, acoustic sensing detects leaks by capturing the vibrations or sound waves generated by high-pressure fluid escaping from the pipeline and converts these signals into electronic data to analyze the leak's size and location [2]. Fiber optic cables, installed along pipeline lengths, change their transmission characteristics when exposed to liquid. Specially designed hydrophones then detect and analyze these signals to pinpoint the leak's location [3]. Infrared thermography identifies abnormal temperature variations in the pipeline environment through real-time thermal detection [4]. While these methods can achieve high accuracy, they often involve high operational costs, complex maintenance, and limited detection ranges, making them less practical for large-scale pipeline systems.

On the other hand, software-based methods rely on models or algorithms to detect and locate leaks using sensor data (e.g., pressure, flow rate, and temperature).

These methods generally require fewer sensors, resulting in lower maintenance costs. Software-based techniques can be divided into two main categories: model-based and data-driven methods. Model-based approaches use theoretical formulas from fluid dynamics to compare computed and measured values, where significant deviations indicate leaks. An example is the Real-Time Transient Model (RTTM) [5, 6]. Data-driven methods include support vector machines [7], convolutional neural networks [8], and wavelet transforms [9], leveraging machine learning, deep learning, anomaly detection, and time series analysis to learn normal system behavior from historical data and detect anomalies indicative of leaks. Meanwhile, purely data-driven (model-based) approaches often require extensive training data or fail to ensure physical consistency in dynamically changing environments. While effective in handling high-dimensional, nonlinear systems, data-driven methods often lack physical interpretability.

This study aims to develop a Physics-Informed Neural Network (PINN) that combines physical knowledge with data-driven approaches for accurate pipeline leakage detection. PINN, introduced by Raissi et al. [10], consists of a neural network for data fitting and a loss function incorporating physical information. By minimizing both prediction and PDE losses, PINN ensures that the model aligns with real-world data and adheres to physical constraints, offering interpretability and reliability. Automatic differentiation (AD) enables efficient computation of partial derivatives during training. In this study, the Navier-Stokes equations, describing fluid pressure and flow variations in the pipeline, are incorporated into the PINN loss function, capturing pipeline dynamics through continuity and momentum equations to enhance leakage detection accuracy. By utilizing only upstream and downstream flow and pressure sensors, our approach reduces operational costs by eliminating the need for expensive hardware, requires minimal maintenance compared to traditional sensor-intensive techniques, enhances long-range leak detection by leveraging both physical knowledge and machine learning capabilities, and ensures physical interpretability, improving reliability in real-world applications.

METHODOLOGY

The schematic diagram of the pipeline system is illustrated in Figure 1. This underground pipeline, designed to transport liquid fluids, extends over a total length of 39.68 kilometers.

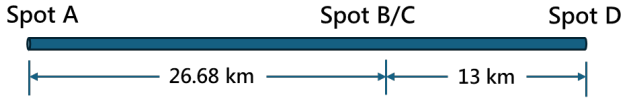

Figure 1. Schematic diagram of the pipeline system.

To achieve the goal of leak detection, it is essential to establish an effective leak detection process, with the logic depicted in Figure 2. By training a PINN model using datasets from normal operating conditions, the model is designed to describe the fluid dynamics within a leak-free pipeline. When the input data provided to the model reflects flow rates and pressures under leak conditions, a statistically significant error arises between the model's predictions and the actual values. Therefore, leak occurrence can be detected by applying statistical control to the prediction errors.

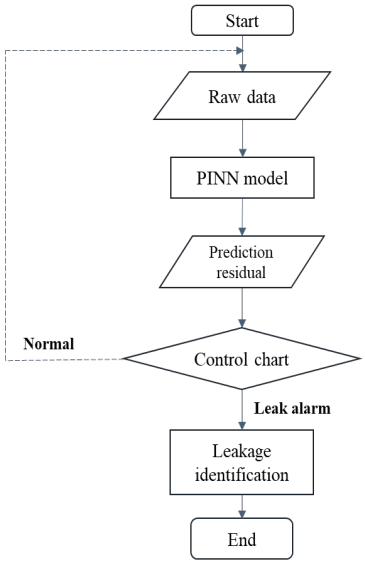

Figure 2. Flow chart for leak detection.

Physics-informed neural networks

To develop a PINN-based surrogate model for predicting pipeline transport variations, it is essential to identify an appropriate physical model and embed it into the loss function to ensure the model adheres to expected physical principles. Considering that the target of this model is to predict the transport behavior of the fluid, the one-dimensional Navier-Stokes equations are employed as the governing equations. These equations describe the motion of fluids (liquids or gases) based on Newton's second law of motion and are derived from the principles of mass conservation and momentum conservation for fluids. The equations are described as follows[11]:

$$\frac{\partial \rho}{\partial t} + \frac{\partial \rho u}{\partial x} = 0 \tag{1}$$

$$\rho \frac{\partial u}{\partial t} + \rho u \frac{\partial u}{\partial x} + \frac{\partial P}{\partial x} + \frac{f \rho u |u|}{2D} = 0 \tag{2}$$

where u is the cross-section averaged velocity (m/s), ρ the density (kg/m^3), p pressure (Pa), f (dimensionless) the Darcy friction factor, D (m) pipe diameter.

To establish a PINN model for flow states governed by partial differential equations (PDEs), a deep neural

network (DNN) is first constructed as a surrogate model for solving PDEs. In this study, a feedforward neural network (FFNN) is employed due to its simplicity, which is sufficient for most PDE problems. Considering that the objective at this stage is to identify abnormal transport states, the model must be capable of recognizing the transport state over continuous time under different initial conditions. To enhance the generalization performance of the model, the input layer includes position x, time t, initial flow velocity condition u_{IC}, upstream-downstream pressure difference ΔP, and pressure P. The model predicts the output $\hat{u}(x, t, u_{IC}, \Delta P, P, \theta)$, where θ is a vector containing all weights and biases within the neural network, which are optimized using the gradient descent algorithm during backpropagation. Variables with a hat denote those computed through the neural network. Furthermore, considering that this case involves an inverse problem, the term f in the equations is treated as a model parameter to be estimated.

Ensuring the model's outputs satisfy both real-world data and the physical laws represented by equations (1) and (2) requires minimizing the residuals g_1 and g_2 as specified in Figure 3. Smaller residual values indicate a high degree of conformity to the underlying physical principles. The time and spatial derivatives are computed using the AD method [12], which eliminates the need for discretizing the computational domain to solve partial differential equations. This approach differs from traditional methods such as the method of characteristics or finite difference methods.

A notable feature of the PINN model is the setup of collocation points, $\{x_f, t_f, u_{ICf}, \Delta P_f, P_f\}$. These points are strategically chosen across different spatial and temporal locations to evaluate the loss based on the problem's requirements. Incorporating these collocation points into the PINN model allows for the calculation of residuals g_1 and g_2 within the loss function, ensuring that the outputs at these points align with the governing partial differential equations. This enables the neural network to leverage physical knowledge while minimizing the reliance on large datasets of real-world data.

The architecture of the model, as illustrated in Figure 3, consists of an input layer that includes position, time, initial flow velocity condition, upstream-downstream pressure difference, and pressure. The output layer predicts flow velocity. The loss function comprises two components: the error between real-world data and predicted values, and the residual error from the governing partial differential equations. In this study, the weights of these components are set as α = 1 and β = 0.1.

The model configuration includes eight hidden layers, each containing 20 neurons. The activation function tanh introduces nonlinearity. For optimization, a two-stage strategy is applied: initial training is performed using the Adam optimizer, followed by fine-tuning with the L-BFGS-B optimizer to complete the training process.

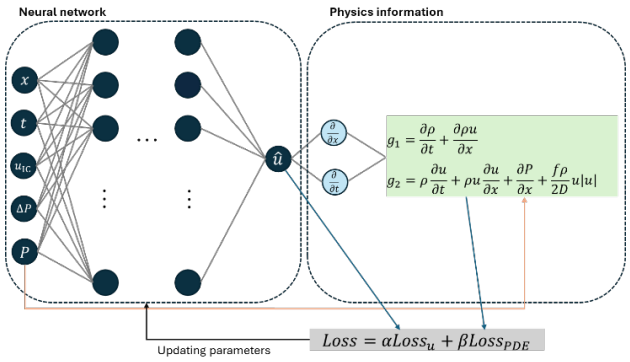

Figure 3. Model structure for PINN.

Dataset splitting and performance evaluation

The dataset used in this study comprises transport data with varying flow rates collected over multiple days of continuous operation. This approach ensures sufficient training of the model and enhances its accuracy and robustness in anomaly detection. The dataset is segmented into fixed windows of 1000 consecutive seconds to create distinct samples, ensuring consistency and precision in the analysis. These samples are further divided into training, validation, and testing sets.

To evaluate the predictive accuracy of the model, the coefficient of determination (R^2) and the root mean square error (RMSE) are adopted as evaluation metrics. The R^2 measures the model's ability to explain the variance in the data. A value closer to 1 indicates that the model more accurately captures the variability in the data, demonstrating stronger explanatory power. RMSE, on the other hand, quantifies the difference between predicted and actual values in terms of absolute error, providing a direct measurement of the model's accuracy. Lower RMSE values signify that the model's predictions are closer to the actual values, offering a quantitative reflection of the model's predictive performance. By combining these two metrics, the model's performance can be effectively compared, providing a basis for further refinement and improvement.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{3}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{4}$$

RESULTS AND DISCUSSIONS

For confidentiality purposes, the flow rate data of the fluid has been normalized. After training the PINN model, the prediction results are shown in Figure 4, which displays the diagonal plots of the actual versus predicted values. In this part, the upstream and downstream flow

rate is separately denoted as Q1 and Q4. The model demonstrates excellent prediction performance for both upstream and downstream flow rates across all datasets, with detailed evaluation metrics provided in Table 1.

Since the dataset includes data from multiple days, variations in transportation conditions or operational states across different days result in the diagonal plots being distributed across different ranges, with each cluster representing data from a specific period. It is important to note that, as the dataset was split into six groups, the calculation of the R^2 value may be biased. Specifically, the denominator of R^2, i.e., the total sum of squares, is relatively large, making the R^2 value closer to 1, which could potentially mislead the results. Therefore, relying solely on R^2 to evaluate the model's performance is not entirely accurate. Instead, RMSE provides a more realistic measure of the model's effectiveness. The RMSE values for Q1 are 0.0005, 0.0007, and 0.0004 for the training, validation, and testing datasets, respectively. Similarly, the RMSE values for Q4 are 0.0007, 0.0005, and 0.0005 for the same datasets. These results indicate that the model effectively minimizes prediction errors and achieves an exceptional level of performance.

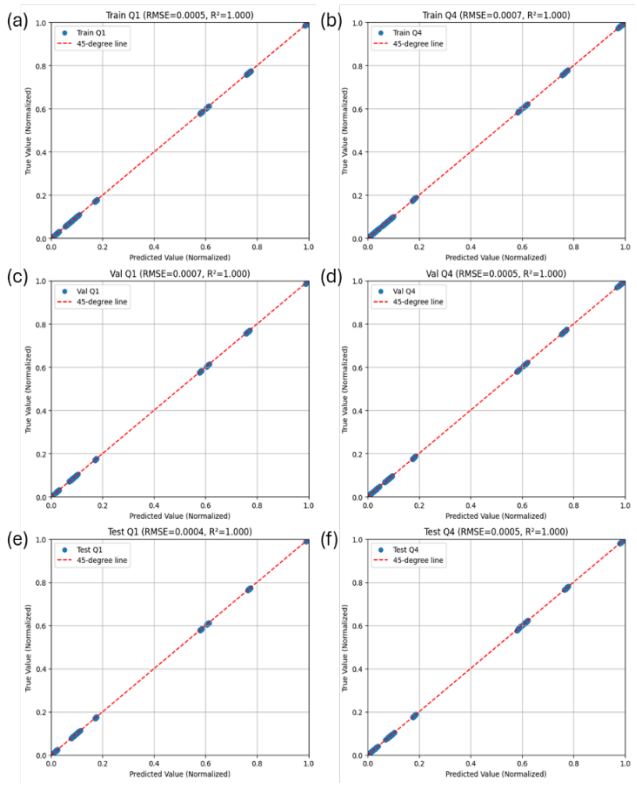

Figure 4. Predicted vs. actual values: (a) upstream flow in the train set, (b) downstream flow in the train set, (c) upstream flow in the valid set, (d) downstream flow in the valid set, (e) upstream flow in the test set, and (f) downstream flow in the test set.

Table 1: Training, validation, and test performance.

Dataset	R^2(Q1/Q4)	RMSE(Q1/Q4)
Train	1.000/1.000	0.0005/0.0007
Valid	1.000/1.000	0.0007/0.0005
Test	1.000/1.000	0.0004/0.0005

To establish a statistical process control (SPC) chart for anomaly detection, this study selects the validation set for residual sampling. The method involves calculating a range that corresponds to 99.73% of the data distribution, analogous to the 6σ range commonly associated with a normal distribution. This range facilitates subsequent anomaly detection and management. The sampling results are shown in Figures 5 and 6. The residual range of upstream flow lies between -0.0227 and 0.0241, while the residual range of downstream flow is between -0.0272 and 0.0306. These ranges serve as benchmarks for subsequent statistical control, used to detect anomalies based on the model's prediction results. If the residuals of the prediction exceed these ranges, they are considered anomalies, triggering alarms.

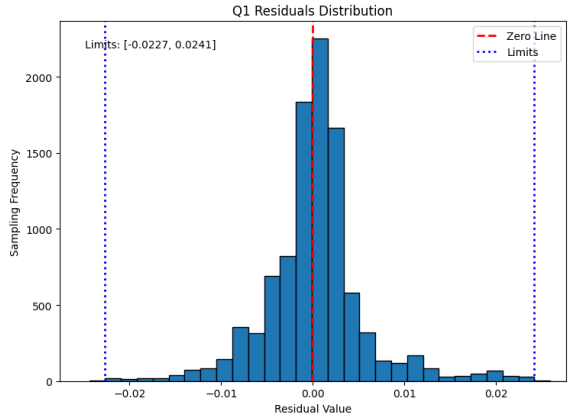

Figure 5. Histograms and estmated control limits of the upstream prediction residuals.

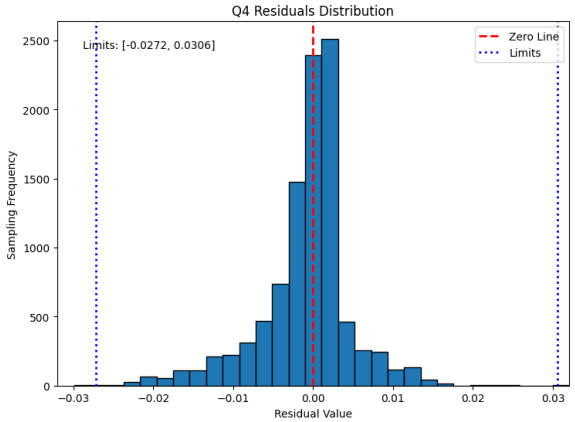

Figure 6. Histograms and estmated control limits of the downstream prediction residuals.

Initially, SPC was performed on normal data that the

model had not seen before. Figures 7 and 8 illustrate the monitoring results for upstream and downstream data, respectively. In the control charts, the x-axis represents time, and the y-axis represents the model's prediction residuals. The blue line indicates the actual residual results, while the red dashed lines represent the upper and lower control limits.

From the figures, it is evident that within the selected 1000-second range, the results show that all normal data fall within the in-control range. Additionally, the actual flow rate prediction results over time are shown in Figure 9. For the upstream and downstream flow data, the R^2 values reached 1.000 and 0.997, respectively. This demonstrates that the model exhibits high accuracy and reliability in predicting under normal transportation conditions.

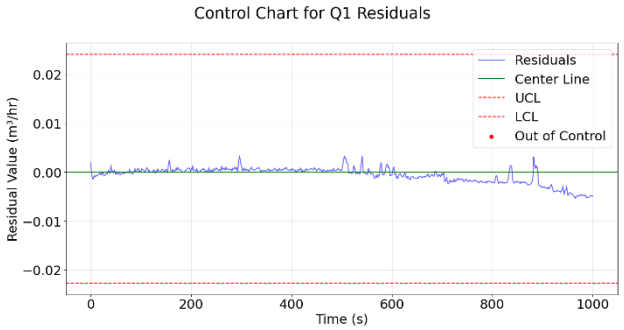

Figure 7. Control chart of upstream residuals under normal transportation conditions.

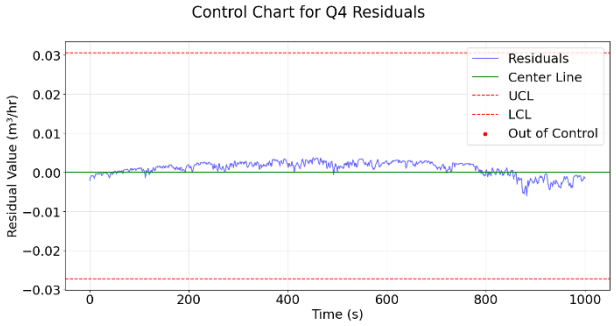

Figure 8. Control chart of downstream residuals under normal transportation conditions.

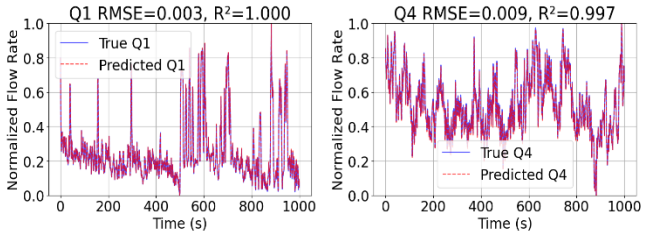

Figure 9. Comparison of predicted and true flow rates under normal transportation conditions at different positions in the pipeline system. (a) Flow rate at upstream, and (b) Flow rate at downstream

In the anomaly detection analysis, the statistical control results shown in Figures 10 and 11 clearly reveal abnormal conditions. The upstream data exceed the upper control limit, while the downstream data fall below the lower control limit. A significant number of data points lie outside the control limits, indicating notable differences compared to normal data and suggesting potential anomalies during this time period. Using the control chart, these leakage features can be identified and extracted. Through the above experiments, this study demonstrates the significant impact of leakage events on the distribution of upstream and downstream pressure and flow rate data. The model accurately captures the abnormal behavior during leakage periods.

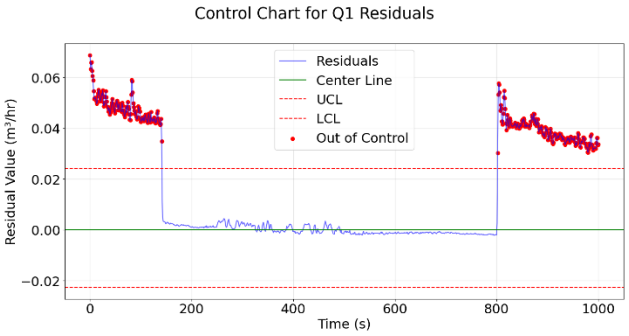

Figure 10. Control chart of upstream residuals under leakage conditions.

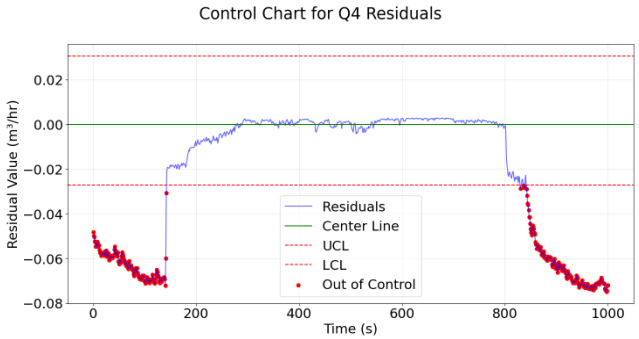

Figure 11. Control chart of downstream residuals under leakage conditions.

CONCLUSIONS

This study developed a pipeline leakage detection system based on Physics-Informed Neural Networks (PINNs), integrating the Navier-Stokes equations of fluid dynamics with deep learning techniques to achieve effective monitoring of pipeline systems. The results demonstrate that the model achieved R^2 values close to 1.000. Furthermore, the RMSE values were 0.0005/0.0007 for the training dataset, 0.0007/0.0005 for the validation dataset, and 0.0004/0.0005 for the testing dataset, confirming the high accuracy of its pre-

dictions. Statistical process control chart based on residual analysis, enabling effective differentiation between normal and abnormal states. For normal condition monitoring, the model exhibited strong predictive capability, with residuals from the training, validation, and testing sets remaining within the control limits. For anomaly detection, the model successfully identified the occurrence of leaks, with a significant number of data points exceeding the control limits. By combining the interpretability of physical models with the predictive power of machine learning, this study provides an innovative solution for industrial pipeline safety monitoring, offering a valuable reference for future pipeline system surveillance and diagnosis.

REFERENCES

1. Yuan J, Mao W, Hu C, Zheng J, Zheng D, and Yang Y. Leak detection and localization techniques in oil and gas pipeline: A bibliometric and systematic review. *Eng Fail Anal* 146:107060 (2023) https://doi.org/10.1016/j.engfailanal.2023.107060
2. Li Z, Zhang H, Tan D, Chen X, and Lei H. A novel acoustic emission detection module for leakage recognition in a gas pipeline valve. *Process Saf Environ Prot* 105:32-40 (2017) https://doi.org/10.1016/j.psep.2016.10.005
3. Tanimola F, and Hill D. Distributed fibre optic sensors for pipeline protection. *J Nat Gas Sci Eng* 1(4-5):134-143 (2009) https://doi.org/10.1016/j.jngse.2009.08.002
4. Manekiya MH, and Arulmozhivarman P. Leakage detection and estimation using IR thermography. *In: 2016 International Conference on Communication and Signal Processing (ICCSP)*. *IEEE*, pp. 1516-1519 (2016) https://doi.org/10.1109/ICCSP.2016.7754411
5. Malekpour A, and She Y. Real-time leak detection in oil pipelines using an Inverse Transient Analysis model. *J Loss Prev Process Ind* 70:104411 (2021) https://doi.org/10.1016/j.jlp.2021.104411
6. Zheng X et al. Leak detection of long-distance district heating pipeline: A hydraulic transient model-based approach. *Energy* 237:121604 (2021) https://doi.org/10.1016/j.energy.2021.121604
7. Qu Z, Feng H, Zeng Z, Zhuge J, and Jin S. A SVM-based pipeline leakage detection and pre-warning system. *Measurement* 43(4):513-519 (2010) https://doi.org/10.1016/j.measurement.2009.12.022
8. Zhou M, Pan Z, Liu Y, Zhang Q, Cai Y, and Pan H. Leak detection and location based on ISLMD and CNN in a pipeline. *IEEE Access* 7:30457-30464 (2019) https://doi.org/10.1109/ACCESS.2019.2902711
9. Xiao R, Hu Q, and Li J. Leak detection of gas pipelines using acoustic signals based on wavelet transform and Support Vector Machine. *Measurement* 146:479-489 (2019) https://doi.org/10.1016/j.measurement.2019.06.050
10. Raissi M, Perdikaris P, and Karniadakis GE. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J Comput Phys* 378:686-707 (2019) https://doi.org/10.1016/j.jcp.2018.10.045
11. Barnard A, Hunt W, Timlake W, and Varley E. A theory of fluid flow in compliant tubes. *Biophys J* 6(6):717-724 (1966) https://doi.org/10.1016/S0006-3495(66)86690-0
12. Paszke A et al. Automatic differentiation in PyTorch. *In: 31st Conference on Neural Information Processing Systems (NeurIPS)*. (2017)