Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 47 No. 01 (2026)

Advances in Pipeline Leak Detection: A Review of Computational Fluid Dynamics, Machine Learning, and Hybrid Diagnostic Frameworks

Koyndrik Bhattacharjee [1,2*], Pronab Roy [1]

[1] Department of Civil Engineering, National Institute of Technology Durgapur, Durgapur 713209, West Bengal, India

[2] Department of Civil Engineering, Dr. B. C. Roy Engineering College, Durgapur 713206, West Bengal, India

** Corresponding Author: koyndrik@gmail.com*

ORCID: K. Bhattacharjee – 0009-0002-5145-4275 | P. Roy – 0000-0001-9042-9369

Abstract

Pipeline leakage is one of the current economically important operational issues in the water distribution systems, petroleum and chemical processing and pharmaceutical fluid procedures networks in the world. Over the past two decades the technology of the leak detection has been transformed fundamentally- beginning with manual scrutiny and primitive pressure assessments to considerably more intricate sensor-based, signal-driven, and the most current concentrating on hybrid-calculation-machine learning diagnostic schemes. The given review critically reviews the current state of the field of pipeline leak detection and particularly how the combination of computational fluid dynamics (CFD) simulation and supervised and unsupervised machine learning can take place. The paper in question categorizes the previous detection methodologies in terms of hardware-based, software-based and hybrid ML-CFD and evaluates the capability, constraints and practicable restrictions of its application with regards to synthesis of over thirty representative publications published in the period between ninth year 1987 and 2025. Much attention is specifically paid to such emergent themes as: Science (i) used physics-consistency of generating synthetic training data via physics simulation through COMSOL Multiphysics and ANSYS based models; (ii) the relative usefulness of classical ML algorithms (Random Forest, Support Vector machine, Ridge Regression, Gradient Boosting) versus deep learning models (CNN, LSTM, hybrid CNNLSTM); (iii) how pre-electra-steady-state single-audioid models can be applied to the problem of variability; variability (iv) the treatment of This review has highlighted the presence of such unanswered challenges as the pre-eminence of steady-state single-fluid research, lack of experimental validation, and lack of standardized benchmark datasets, and suggests a systematic research agenda to fully autonomous, multi-fluid, real-time pipeline integrity management systems.

Keywords: pipeline leak detection; machine learning; computational fluid dynamics; review; Random Forest; deep learning; multi-fluid; multi-leak; COMSOL simulation; acoustic emission; predictive maintenance.

1. Introduction

The pipeline systems also play the significant role of dispatching water, petroleum, natural gas, and industrial chemicals all across the globe that spans over three point five millennium kilometers of transmission and distribution system. The security and efficient operation of such networks mechanisms are the core focus of the overall population, energy security, and industrial output [1, 2]. However, corrosion, mechanical damage, fatigue cracking, joint failures and environmental stressors are all long term significant sources of threat to pipeline integrity. The tiny unidentified leakages might evolve into catastrophic loss, in the form of economic loss of billions of dollars annually besides increasing the harmful environmental pollution [3, 4].

The progress of methodological history of leak detectors has undergone considerable development throughout the past 4 decades beginning with the initial-generation leak detection instruments of manual and visual inspection, the second-generation innovation of sensor-based continuous production to the current fourth-generation instrument of continuously combining physics-based numerical simulation instruments and data-based machine learning [1, 5, 6]. With this liberalized sophistication, however, there exist some gaps in knowledge that are critical. Even published literature is limited to to single-fluid and single-leak cases in large numbers. Limited controlled benchmarking of the various algorithms to controlled conditions is done. And the synthesis of physical intuition at work through simulation with the predictions achieved by the nation ML remains a research frontier that is yet to be fully pursued [7, 8, 9].

The review paper is a blanket assessment of the corresponding field of discussion, the categorization of detection methodologies, and a critical analysis of their scope of benefits and limitations, which is assessed by discussing exemplary published works, delineating current tendencies, and gaps in the existing research, and offers the research directions in the future. As opposed to systematic literature reviews based on formal search guidelines, the paper will be presented on a narrative review framework according to which, it is stated that their focus should be on the synthesis of themes and the critical evaluation of the methodological paradigms rather than the exhaustive enumeration of the existing studies. The paper can be divided into ten sections that comprise of detection taxonomy (Section 2), CFD simulation methods (Section 3), analytical leakage modeling (Section 4), ML algorithms (Section 5), deep learning advances (Section 6), multi-fluid and multi-leak extensions (Section 7), post-processing and deployment (Section 8), gaps in research and future directions (Section 9), and conclusions (Section 10).

2. Taxonomy of Pipeline Leak Detection Methods

Techniques of pipeline leak detection can be roughly divided into three major groups: hardware-based (external) techniques, software-based (internal/computational) techniques, and a hybrid class of methods that combine both techniques. The different categories consist of different technological families, which have their respective strengths, weaknesses, and area of use. This is the systematic summary of what each category entails using sample studies to highlight major strengths and limitations.

2.1 Hardware-Based Detection Methods

We have the Hardware-based approaches which use physical sensors placed along the pipeline or inside the pipeline in order to detect evidence of leakage events. The high-frequency stress waves occurring as stressed fluid escapes through an opening of a leakage are detected by acoustic emission (AE) technology. Ullah et al. [10] showed that the statistical characteristics of AE signals such as base, skewness and root mean square amplitude can be used as useful inputs to a trained classifier to detect pinhole-scale leakage. The Butterfield et al. [11] have developed correlations between the vibro-acoustic signature and the leak flow rate of polyethylene pipes. Nevertheless, AE methods have severe signal attenuation in long pipes and are prone to ambient noise in factories [12].

The most commonly used industrial detection technology is pressure-based monitoring which detects leaks based on the characteristics of pressure transients that travel out of the source. The use of negative pressure wave (NPW) detection and real-time transient modeling (RTTM) has been used widely in oil and gas pipelines [1, 5]. Field testing of NPW systems has been carried out by Siebenaler et al. [13] which confirms systems to be effective with large leakages but records low sensitivity to small leaks generating pressure changes within the operating noise band. Fiber-optic distributed temperature sensing (DTS) is also another hardware method, and it has been shown to detect temperature changes caused by the leakage of fluids [14]. Flow-based techniques that compare the inlet and outlet mass balance give an easy detection principle and cannot localize the location of leak places and are restricted by meter accuracy [5, 15].

Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 47 No. 01 (2026)

2.2 Software-Based and Model-Driven Methods

Software based techniques include mass balance analysis, Kalman filtering, parameter identification and inverse transient analysis (ITA). Billmann and Isermann [15] were the first to specify model-based leak detection by identifying the parameters of pipeline hydraulic models. Torres et al. [16] have reviewed the uses of Kalman filter in the diagnosis of pipeline leakages, and they have proved their benefits in state estimation in case of noisy measurements. Malekpour and She [17] designed real-time based ITA systems based on genetic algorithm optimization attaining strong leak localization in simulated SCADA noise. The study conducted by Zhang et al. [18] suggested a better particle swarm optimization in the inverse hydro macro models with high anti-noise capabilities and low false alarms.

Although software methods are mathematically elegant, their reliance on valid hydraulic models, calibration error sensitivity, and cost, which makes them computationally expensive, are still a weakness [1, 3, 16].

2.3 Hybrid ML-CFD Methods

Recent advances in leak detection technology 3 involve combining physics-based simulation of CFD with machine learning to form hybrid diagnostic models that simultaneously enjoy both the physical intelligibility of mathematical models and the pattern recognition and generalization of ML methods. The effectiveness of this method has been proved in a number of studies: Bhattacharjee and Roy [7] employed polynomial regression together with the Random Forest ensemble learning in the context of synthetic data obtained via COMSOL and obtained close accuracy in prediction (R^2 = 0.998 when predicting leak size). The Gao et al. [19] suggested the PSO-optimized SVM to detect oil-gas pipeline leakages by using the negative pressure wave characteristics. Most recently, there are proposals of multi-fluid [8] and multi-leak [9] versions of the hybrid framework such that much greater applicability is achieved. In the body of this paper, each of these aspects of this hybrid paradigm are discussed separately.

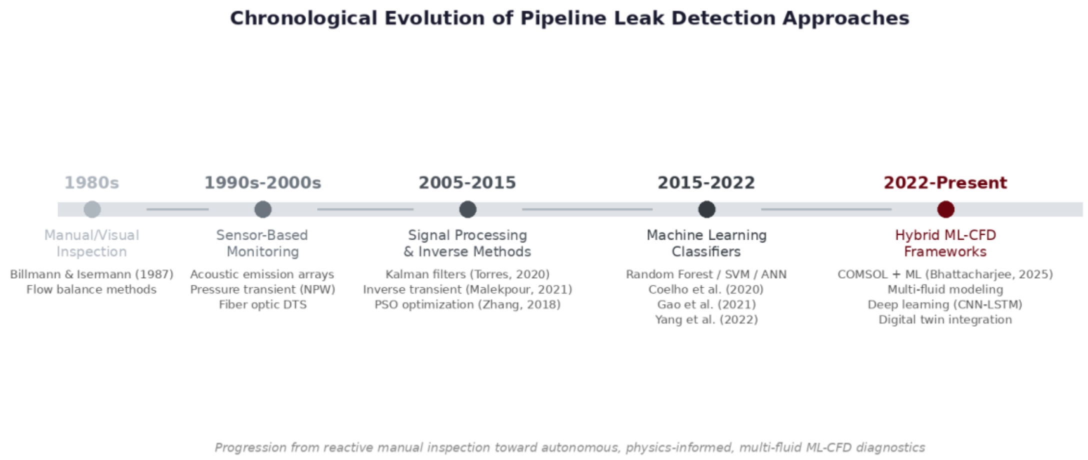

Figure 1. Chronological evolution of pipeline leak detection methodologies, illustrating the progressive transition from manual inspection (1980s) through sensor networks and signal processing to contemporary hybrid ML-CFD frameworks. Representative studies are noted at each evolutionary stage.

Table 1. Comparative assessment of leak detection method categories: capabilities, limitations, and representative applications

Method Category	**Representative Technologies**	**Principal Strengths**	**Critical Limitations**

Hardware-Based	AE sensors, pressure transducers, fiber optic DTS, flow meters, pigging/ILI [1,10,11,13]	Real-time capable; well-established industrial standards; direct physical measurement	Signal attenuation; background noise; high deployment cost; cannot localize small leaks
Software-Based	Kalman filter, ITA, RTTM, mass balance, parameter ID [15,16,17,18]	No additional hardware; analytical framework; optimization-based localization	Model calibration dependency; false alarms from transients; limited real-time scalability
Hybrid ML-CFD	RF + COMSOL, SVM + CFD, CNN-LSTM, ensemble models [7,8,9,19,20]	Nonlinear pattern capture; multi-fluid generalization; physics-informed; automated feature learning	Requires simulation or experimental data; limited field validation; computational training cost

3. Role of CFD Simulation in Leak Detection Research

Computational fluid dynamics has become an enabler in current leakage research, playing two key functions, (i) as a research instrument of deepening insight into the underlying pressure-velocity relationships in leakage flow, and (ii) as a synthetic data generator to train and validate machine learning models where large-scale experimental data are unavailable.

3.1 Simulation Platforms and Governing Physics

The simulation platforms used most active in the literature review are COMSOL Multiphysics, ANSYS Fluent, and OpenFOAM. COMSOL has especially been popular in laminar flow interface and the versatile multi-physics coupling features [7, 8, 9]. ANSYS Fluent is used to conduct turbulence modeling, as well as large-scale pipeline optimization simulations [21, 22]. The controlling physics in each scenario entails resolving of the steady or changeable Navier-Stokes equations on incompressible Newtonian fluids that describe the pressure drop, velocity redistribution and jetting development, which characterize the flowings of leaks [7, 8].

3.2 Simulation-Based Dataset Generation

One of the most important benefits of CFD-based methods is the possibility to produce large and parametrically controlled datasets capable of systematically sampling the pipeline geometry, leak settings, flow conditions, and fluid properties space. Bhattacharjee and Roy [7] used COMSOL to create 500 leak scenarios by changing velocity, pressure, leak size, and position position to create a synthetic training set with both the activities of a basic regression and a more complex random forest model. Oseni et al. [23] utilized the finite element CFD to compare the pressure and velocity fields to use as leak indicators in crude oil pipelines and found that the pressure fields are more effective as leak indicators compared to the velocity field in turbulent environments. Manshoor et al. [24] modeled the leak flow properties in oil and gas pipeline networks and obtained the confirmation that pressure monotonically decreases in the existence of leakages.

Although the benefits of simulation-based datasets are parametric control and reproducibility, they are an approximation compared to the real-world situation that can fail to represent material heterogeneity, sensor noise, or variability in the environment. It is thus understood in the field that experimental validation is a necessary supplement to model development by simulation [7, 8, 25].

3.3 Multi-Fluid Simulation: An Emerging Direction

It has also brought to language a notable methodological development, namely the generalization of single-fluid models to multi-fluid models. Simulating four industrially competing fluids, solutions of water, ethanol, diesel, and glycerin, at the same pipeline geometry, Bhattacharjee and Roy [8] found that the distributions of pressure were highly sensitive to the density-viscosity interaction even in situations where the boundary conditions are assumed constant. Tavares et al. [26] theorized the oil-water-gas multiphase leakage and found that phase interactions result in multiple transient responses and that the leak pressure signature is radically changed. These multi-fluid models deal with one fatal flaw of the literature, in which the vast majority of models assume water as the working fluid but the fluids in industrial pipelines are multi-fluid with many varieties of fluids passing through [8].

4. Analytical Leakage Equations and Physical Modeling

Some of the studies have used the curve-fitting regressions to extract explicit analytical expressions between the properties of leaks to quantifiable hydraulic parameters. Bhattacharjee and Roy [7] have formulated two-degree polynomials of leak size and leak position as functions of velocity and pressure giving models that are easily interpretable whose coefficients hold a relative value that indicates the direct impact of any parameter. The coefficient analysis showed that velocity prevails in prediction of leak size both linearly and quadratically, and the pressure has a relatively low opposing effect.

One of the extensions was the re-parameterization into definite dimensionless groups of these equations that included fluid density [8]. The four equations can be applied to fluids of varying densityviscosity behavior by simply scaled by velocity V/V in and pressure P/ (rhoV in 2). The theoretical framework of the underlying hydraulic model can be used without changes. This is a density-augmented formulation, which is a contribution to multi-fluid leak characterization that allows the same analytical framework to be used in water distribution applications, fuel transport, and chemical processing applications [8].

Physical modeling is an alternative to machine learning since it can make transparent predictions grounded on equations that can be tested in isolation against fluid mechanics theory. A number of authors have proposed that the interpretation model and high-accuracy ensemble ML hybrid body offer the most promising future avenue towards the development of diagnostics systems to meet both transparency-associated engineering and operational representation criteria [7, 8, 27].

5. Machine Learning Algorithms for Leak Detection: A Critical Evaluation

Machine learning applications to the detection of pipeline leaks have grown intensively since the early 2010s, including classification problems (leak vs. no-leak) and regression problems (size, position, pressure, and flow rate of a leak). This segment is a critical overview of the major algorithm family implementations that have been put into practice based on relative benchmarking studies to determine their comparative strengths and shortcomings.

5.1 Ensemble Methods: Random Forest and Gradient Boosting

Random Forest has become one of the most successful algorithms that is recurrently successful on the reviewed literature. Based on the findings of Coelho et al. [20], the Random Forest, decision trees, neural networks and SVMs have been compared in detecting water leaks using wireless sensor at maximum accuracy of 75.75% in real life setting. Random Forest has shown almost perfect R2 ([?]0.998) in leak prediction in simulation-driven conditions [7] and the best cross-fluid stability of all the algorithms in the benchmark [8]. This characteristic noise resilience, embedded metrics of feature importance and overfitting resistance make it especially appropriate in applications where users are interested in the pipeline using sensor data as an input and physical interpretability as an output [7, 8, 27].

Gradient Boosting Regression (GBR) has also performed well, and its sequential error-correction algorithm has been shown to give high-accuracy in nonlinear interactions between pressure and velocity in many fluids [8]. GRB however, demands more critical hyperparameter optimization than random forest to prevent overfitting especially when using less simulation data [8, 28].

5.2 Support Vector Machines

Another popular method of leaks detection with Support Vector Machines and its derivatives (SVM, One-Class SVM, SVR) have been used in classification and regression. Gao et al. [19] also came up with PSO-optimized SVM to detect oil-gas pipeline and showed that it outperforms existing models with optimization of kernel parameters. Anomaly detection One-Class SVM has been used in detection of anomalies described in the situation where the number of leak events is small compared to the normal operation [29]. Nevertheless, it has been shown that multi-fluid benchmarking allows to notice a sharp deterioration in the performance of the SVR in case of fluid property variations with the maximum error variances among all the tested algorithms (stability index = 1.12x106) [8]. This instability indicates that there may be a bad fit of kernel-based methods to multi-fluid deployment unless the per-fluid calibration is significant.

5.3 Linear and Regularized Models

A regularized linear model Ridge Regression has been actively tested with singe-fluid water systems showing the best results with MAE (lowerest) (157.01 Pa) in a single benchmarking study [28]. But its linearization form essentially restricts its capacity to include the nonlinear densityviscosity interactions of multi-fluid leakage and prediction error in the latter escalates significantly between water and glycerin [8]. These results indicate that linear models could be used as valid baseline comparators or interpretability measurements but should not be trusted to be used as operational predictors in operationally diverse pipeline environments.

5.4 Neural Networks

Multilayer Perceptrons (MLP) have been used with ambivalent success in leak detection. Even though its performance theoretically can be as good as that of any approximation to a function, MLP has always shown the worst performance among those tested in simulated conditions, both in error terms (MAE up to 760 Pa) and in values of R2 (0.68-0.71) in a variety of fluids [8, 28]. This poor performance is explained by the fact that simulation-derived datasets are small, and they are not enough to restrict the huge parameter space of neural network designs and encourage overfitting [7, 28]. These results can be discussed in connection with the larger literature about ML that suggests that the average ensemble approach is generally more successful when compared to neural networks with small-to-moderate datasets.

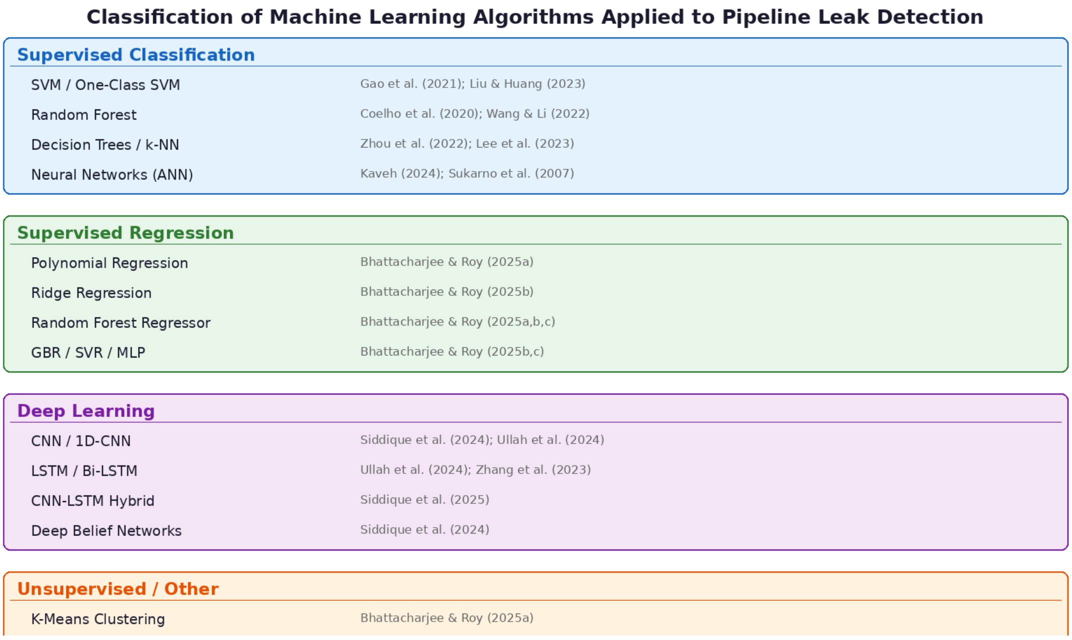

Figure 2. Classification of machine learning algorithms applied to pipeline leak detection across four methodological families: supervised classification, supervised regression, deep learning, and unsupervised/optimization approaches. Representative studies are cited for each algorithm.

6. Deep Learning and Emerging Computational Approaches

Deep learning architectures are becoming more popular with pipeline leak detection, especially acoustic emission signal detection and image leak classification. Siddique et al. [30] proposed a deep belief network (DBN) fine-tuned using genetic algorithms and combined with least-squares SVM in CWT scalogram-based leak detection with high levels of classification accuracy. Ullah et al. [31] introduced the CNN-LSTM hybrid framework, which transforms AE signal to scalogram image, and highly detects the real-world pipeline data with a 99.69 per cent accuracy.

In their study, Xi et al. [32] joined the algorithms of district metering areas with cuckoo search optimization to determine the leaks in the water network of the city through the deep learning neural networks. The shift towards temporal deep learning-based models, such as LSTM and bidirectional LSTM, is indicative of the understanding that the dynamics of the leak evolution is time-dependent and cannot be described fully by the regression-based models that do not change over time [31, 33].

The improvement of the training of ML models with the help of metaheuristic optimization has been implemented. One study conducted by Kaveh and Khavaninzadeh [27] proved that in engineering problems, Enhanced Colliding Bodies Optimization (ECBO) can significantly enhance ANN prediction accuracy, and the results of this research can be applied to leak detection settings. Kaveh [34] also furnished an in-depth discussion on the application of ANN and ML to civil engineering and gave a theoretical basis of the use of the neural network surrogates in a nonlinear system such as hydraulic leak prediction.

Although these developments are optimistic, the pipeline leak detection field has a number of challenges in deep learning methods: large labeled datasets needed to assess the pipeline are challenging to find in the field, computation is expensive and might significantly limit on-edge computing devices where the method should operate, and the problem is less interpretable compared to some ensemble-based methods [30, 31, 34].

7. Multi-Fluid and Multi-Leak Extensions

7.1 Multi-Fluid Modeling Frameworks

The application of leak detection structures in multi-fluid contexts as a continuation of the single-fluid context is a significant step towards viable use in industrial contexts. The overwhelming majority of published literature assumes the use of water as the working fluid, but in reality, pipeline network systems flow fluids which have dramatically different density viscosity properties [8]. The initial systematic comparison of the performance of ML models on the four fluids (water, ethanol, diesel, glycerin) was proposed by Bhattacharjee and Roy [8]) whereby, the ensemble approaches (Random Forest, GBR) yield stable accuracy values with the change of fluid properties, whereas linear (Ridge) and kernel-based (SVR) models lose their performance dramatically. Their inter-fluid stability index (annual variance of MAE between fluids) is a novel quantitative index of how robust a model is to interventions affecting fluids.

Tavares et al. [26] showed based on multiphase simulation that the density and viscosity of the fluid change the transient pressure-velocity caused by leakage fundamentally, to validate the physical essence of fluid-specific modeling. ANN based leak predictors have been created by Sukarno et al. [35], who trained their predictors with synthetic data and varying oil densities and viscosities, and showed that their predictors are successful in predicting the leak rate and location with different fluid types. All these studies confirm that both analytical and ML-based leak detection systems need fluid properties to be inherently added to be cross-domain.

7.2 Multi-Leak Interaction Effects

The application of leak detection structures in multi-fluid contexts as a continuation of the single-fluid context is a significant step towards viable use in industrial contexts. The overwhelming majority of published literature assumes the use of water as the working fluid, but in reality, pipeline network systems flow fluids which have dramatically different density viscosity properties [8]. The initial systematic comparison of the performance of ML models on the four fluids (water, ethanol, diesel, glycerin) was proposed by Bhattacharjee and Roy [8])

whereby, the ensemble approaches (Random Forest, GBR) yield stable accuracy values with the change of fluid properties, whereas linear (Ridge) and kernel-based (SVR) models lose their performance dramatically. Their inter-fluid stability index (annual variance of MAE between fluids) is a novel quantitative index of how robust a model is to interventions affecting fluids.

Tavares et al. [26] showed based on multiphase simulation that the density and viscosity of the fluid change the transient pressure-velocity caused by leakage fundamentally, to validate the physical essence of fluid-specific modeling. ANN based leak predictors have been created by Sukarno et al. [35], who trained their predictors with synthetic data and varying oil densities and viscosities, and showed that their predictors are successful in predicting the leak rate and location with different fluid types. All these studies confirm that both analytical and ML-based leak detection systems need fluid properties to be inherently added to be cross-domain.

Table 2. Comparative summary of representative simulation-driven leak detection studies reviewed

Study	**Fluid(s)**	**Leak Config.**	**ML Models Applied**	**Best Performer**	**Key Contribution**
Bhattacharjee & Roy [7]	Water	Single leak	Polynomial + RF	RF (R^2=0.998)	Dual-model hybrid; K-Means risk clustering
Bhattacharjee & Roy [28]	Water	Single leak	RF, GBR, Ridge, SVR, MLP	Ridge (MAE=157)	5-algorithm benchmark; feature importance analysis
Bhattacharjee & Roy [8]	Water, Ethanol, Diesel, Glycerin	Single leak	RF, GBR, Ridge, SVR, MLP	RF (stability=9.31)	Multi-fluid framework; density-augmented equations; stability index
Bhattacharjee & Roy [9]	Water	2–4 concurrent leaks	Numerical (no ML)	N/A	Multi-leak interaction; spacing effects; diameter influence
Coelho et al. [20]	Water	Real-world	RF, DT, ANN, SVM	RF (75% acc.)	Wireless sensor + ML for real-world WDN
Gao et al. [19]	Oil-gas	NPW signals	PSO-SVM	PSO-SVM	Metaheuristic kernel optimization
Ullah et al. [10]	Water/Gas	Pinhole leaks	NN, DT, RF, k-NN	RF	AE feature extraction with ML classifiers
Siddique et al. [30]	Water	Various sizes	DBN-GA-LSSVM	Hybrid DL	CWT scalogram + deep learning
Oseni et al. [23]	Crude oil	Single leak	FEM + analysis	N/A	Pressure vs. velocity as leak indicators

8. Post-Processing, Deployment, and Engineering Considerations

8.1 Feature Importance and Physical Interpretability

Another unique benefit of the tree-based ensemble methods is that they provide rankings of feature importance i.e.one can check whether the learned relationships are physically consistent. Featuring in various investigations, local velocity at the leak point was found to be the most powerful predictor of pressure anomalies with more than 60-percent of predictive power in Random Forest models, second only to inlet velocity (≈ 20 per cent) and pipe length (≈ 15 per cent) [7, 28]. Pipe size only adds about 5 percent, which proves that pressure signatures during leakage arise out of local rather than global geometric influences. These rankings are consistent with known principles of fluid mechanics, and it has been confidently believed that the ML models are learning real relationships of hydraulics and not artifacts of the dataset at hand [7, 8, 28].

8.2 Unsupervised Clustering for Risk Categorization

According to Bhattacharjee and Roy [7], K-Means clustering was proposed to be conducted as a post-processing step so that continuous regression results could be converted to operationally useful risk classes (low, medium, high severity). The strategy will close the divide between algorithmic forecasting and maintenance decision-making and allow risk-stratification of inspection scheduling and resources distribution. The computer-aided methodology can be extended to wider applicability to leak detection and any predictive maintenance situation where continuous model output has to be discrete-time operational categories [7].

8.3 Deployment Architecture Considerations

To be used operationally, the ML-based leak detectors should be interchangeable with the existing SCADA (Supervisory Control and Data Acquisition) infrastructure and have the ability to meet real-time processes. Due to its parallelizable maturity of tree evaluation and its moderate memory requirements, random forest in particular can be deployed using edge computing platforms with ease [7, 8]. Although deep learning models are more precise to specific tasks, they require more computational resources and cloud-based inference architectures may be necessary. The analysed literature tends to point to the fact that the most efficient deployment framework can consist of ensemble networks to perform real-time edge detection with periodic deep learning-based analysis of complex multi-sensor fusion problems [30, 31, 34].

9. Research Gaps, Open Challenges, and Future Directions

Although the progress in this area is quite impressive and has been recorded in this review, there are still a number of research gaps that characterize the frontier of the field. The key issues that remain unsolved are identified in this section and an organized roadmap in the direction of further research is suggested.

To start with, the prevalence of steady-state laminar flow in simulation-based research is a major limitation since the actual pipelines in real life are subject to transient conditions, turbulent flow regimes, time-varying flow demands and thermal influences. It would be significant to extend ML benchmarking to transient and turbulent simulations to make it more applicable in practice [8, 9]. Second, available literature has practically used synthetic simulation data. Experimental validation on a large (laboratory) and field scale, including real sensor noise, communication latency, ageing hardware (pipes) and/or the variable nature of the environment, would play a critical role in proving operational readiness [7, 8]. Third, the single fluid (water-only) experiments are still conducted although industrial applications have a variety of fluid types. The multi-fluid framework suggested by Bhattacharjee and Roy [8] is a significant first step although their application is extended to non-Newtonian fluids, multi-phase mixtures as well as thermally dissimilar conditions.

Fourth, the dynamic nature of leak development time-dependence, that is, the change in the size, severity, and location of leaks during a period has not been covered by the static regression models that were used in the reviewed studies. Recurrent architectures (LSTM, GRU) and Temporal Convolutional Networks (TCN) provide the opportunity to have prognostic capabilities that forecast a future leak state according to the past development [31, 33]. Fifth, this study has small bundle of reproducibility and cross-study comparison is restricted by the lack of standardized and open-access benchmark datasets to detect pipeline leaks. The benchmark dataset project of Aghashahi et al. [25] is a noble move although efforts significantly larger are required. Sixth, integration of digital twins [34]--sustaining constantly prepared virtual versions of physical pipeline systems] is a radically new chance at predictive maintenance yet to be implemented successful to date in the leak detection system [34]. Lastly, transfer learning methods that allow one to have models that have learned in a specific fluid-geometry setup and apply them to different ones with minimal or no retraining may be incredibly cost saving in terms of deployment.

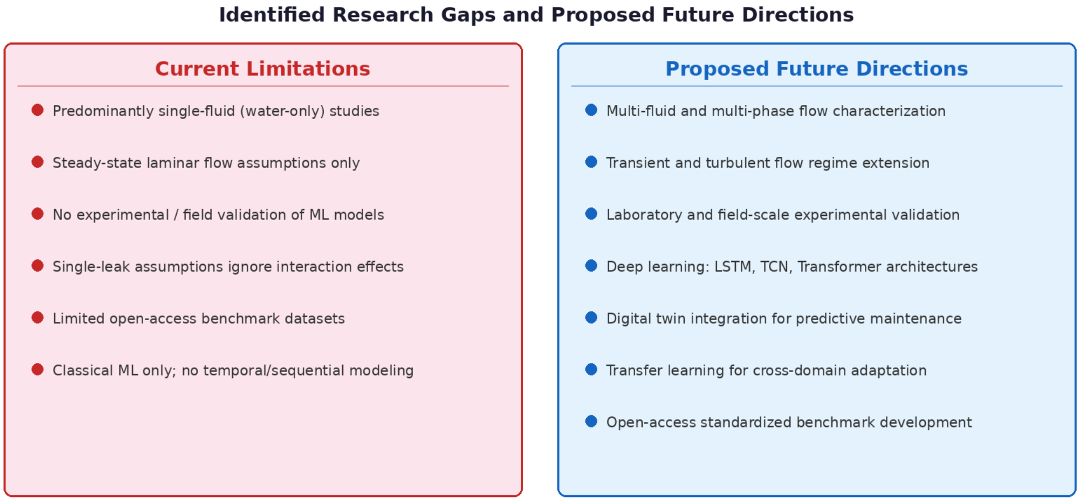

Figure 3. Summary of identified research gaps (left) and proposed future research directions (right), providing a structured roadmap for advancing the field toward autonomous, multi-fluid, real-time pipeline integrity management.

10. Conclusion

The changing face of pipeline leak detection has been critically reviewed as the methodological development of manual inspection and pressure measurement to sensor-based and signal-processing-based methods up to the modern merging of computational fluid dynamics and machine learning. A review of more than thirty representative studies shows that there are some overriding conclusions reached after forty or more years of investigations.

Machine learning models based on ensembles, specifically Random Forest and Gradient Boosting Regression have continuously provided the best predictive accuracy, cross-fluid generalization and noise resilience compared to linear, kernel-based and neural network models in the literature reviewed. Their inbuilt feature importance measures offer vital connection to literalizability which the more detailed architectures tend to miss. The use of CFD simulation and in particular the COMSOL Multiphysics has become an irreversible force in providing physics-consistent synthetic data that allows systematic benchmarking of ML run in different parameter space. The single-fluid gap This is a gap in the single-fluid literature, which the multi-fluid extension of the analytical equations (density-augmented), and cross fluid stability indexing close. Current leakage analysis: It is discovered that due to the effects of coupled and nonlinear interactions of concurrent leakage events, single-leak superposition approaches cannot be used to predict leakage and important special multi-leak diagnostic techniques are needed.

Everything in the future of the field is radical. The areas of potential research that appeared to be the most influential are the optical introduction of the temporal deep learning architecture to learn prognostic leaks, development of the digital twin frameworks to make predictions by the basis of maintenance, and the development of open-ended benchmark data to compare algorithms in an objectively reproducible manner. The interdisciplinary collaboration of the branches of fluid mechanics, machine learning, sensor engineering, and infrastructure systems science will be required over a long period of time to realize the vision of having fully autonomous and multi-fluid real-time operations of managed pipeline integrity.

References

[1] M. A. Adegboye, W. K. Fung, and A. Karnik, "Recent advances in pipeline monitoring and oil leakage detection technologies: Principles and approaches," Sensors, vol. 19, no. 11, p. 2548, Jun. 2019. doi: 10.3390/s19112548.

[2] R. Cramer, D. Shaw, R. Tulalian, P. Angelo, and M. V. Stuijvenberg, "Detecting and correcting pipeline leaks before they become a big problem," Marine Technology Society Journal, vol. 49, no. 1, pp. 31–46, 2015. doi: 10.4031/mtsj.49.1.1.

[3] S. K. Vandrangi, T. A. Lemma, S. M. Mujtaba, and T. N. Ofei, "Developments of leak detection, diagnostics, and prediction algorithms in multiphase flows," J. Loss Prev. Process Ind., vol. 72, Art. no. 104536, 2021. doi: 10.1016/j.jlp.2021.104536.

[4] M. S. El-Abbasy, A. Senouci, T. Zayed, L. Parvizsedghy, and F. Mirahadi, "Unpiggable oil and gas pipeline condition forecasting models," J. Perform. Constr. Facil., vol. 30, no. 1, 2016. doi: 10.1061/(ASCE)CF.1943-5509.0000716.

[5] N. V. S. Korlapati, F. Khan, Q. Noor, S. Mirza, and S. Vaddiraju, "Review and analysis of pipeline leak detection methods," J. Pipeline Sci. Eng., vol. 2, Art. no. 100074, 2022. doi: 10.1016/j.jpse.2022.100074.

[6] Y. Shao, T. Yao, J. Gong, J. Liu, T. Zhang, and T. Yu, "Impact of main pipe flow velocity on leakage and intrusion flow: An experimental study," Water, vol. 11, no. 1, p. 118, 2019. doi: 10.3390/w11010118.

[7] K. Bhattacharjee and P. Roy, "Simulation-driven leak diagnostics in pipelines: Machine learning and curve fitting-based prediction models," Asian J. Civ. Eng., vol. 26, no. 11, pp. 4739–4751, 2025. doi: 10.1007/s42107-025-01453-1.

[8] K. Bhattacharjee and P. Roy, "A multi-fluid machine learning framework for leakage prediction in pressurized pipeline systems," 2025. [Manuscript submitted for publication].

[9] K. Bhattacharjee and P. Roy, "Flow distribution and pressure degradation in multi-leak pipelines," 2025. [Manuscript submitted for publication].

[10] N. Ullah, Z. Ahmed, and J.-M. Kim, "Pipeline leakage detection using acoustic emission and machine learning algorithms," Sensors, vol. 23, no. 6, p. 3226, 2023. doi: 10.3390/s23063226.

[11] J. D. Butterfield, V. Meruane, R. P. Collins, G. Meyers, and S. B. Beck, "Prediction of leak flow rate in plastic water distribution pipes using vibro-acoustic measurements," Struct. Health Monit., vol. 17, no. 4, pp. 959–970, 2017. doi: 10.1177/1475921717723881.

[12] F. Zohora, "Study of pipe leak fluid dynamic characteristics and their influences on acoustic emission generation," Ph.D. dissertation, Queensland Univ. of Technology, 2021. doi: 10.5204/thesis.eprints.207823.

[13] S. Siebenaler, E. J. Tervo, P. Vinh, and C. Lewis, "Field testing of negative-wave leak detection systems," in Proc. IPC 2014, Calgary, Canada, 2014. doi: 10.1115/IPC2014-33557.

[14] M. Shehadeh, A. Sharara, M. Khamis, and H. El-Gamal, "A study of pipeline leakage pattern using CFD," Can. J. Mech. Sci. Eng., vol. 3, no. 3, pp. 98–101, 2012.

[15] L. Billmann and R. Isermann, "Leak detection methods for pipelines," Automatica, vol. 23, no. 3, pp. 381–385, 1987. doi: 10.1016/0005-1098(87)90011-2.

[16] L. Torres, J. Jiménez-Cabas, O. M. González, L. Molina-Espinosa, and F. López-Estrada, "Kalman filters for leak diagnosis in pipelines: Brief history and future research," J. Mar. Sci. Eng., vol. 8, no. 3, p. 173, 2020. doi: 10.3390/jmse8030173.

[17] A. Malekpour and Y. She, "Real-time leak detection in oil pipelines using an inverse transient analysis model," J. Loss Prev. Process Ind., vol. 70, p. 104411, 2021. doi: 10.1016/j.jlp.2021.104411.

[18] H. Zhang, Y. Liang, W. Zhang, N. Xu, Z. Guo, and G. Wu, "Improved PSO-based method for leak detection and localization in liquid pipelines," IEEE Trans. Ind. Inform., vol. 14, no. 7, pp. 3143–3154, 2018. doi: 10.1109/TII.2018.2794987.

[19] J. Gao, Y. Zheng, K. Ni, H. Zhang, B. Hao, and J. Yan, "Research on oil-gas pipeline leakage detection method based on PSO-optimized SVM," J. Phys. Conf. Ser., vol. 2076, no. 1, p. 012004, 2021. doi: 10.1088/1742-6596/2076/1/012004.

[20] J. Alves Coelho, A. Glória, and P. Sebastião, "Precise water leak detection using machine learning and real-time sensor data," IoT, vol. 1, no. 2, pp. 474–493, 2020. doi: 10.3390/iot1020026.

[21] R. Ferraiuolo, F. De Paola, D. Fiorillo, G. Caroppi, and F. Pugliese, "Experimental and numerical assessment of water leakages in a PVC-A pipe," Water, vol. 12, no. 6, p. 1804, 2020. doi: 10.3390/w12061804.

[22] M. Govender, D. Ikegwuoha, and M. Seyam, "A CFD-based study of pipe leakage dynamics and water management impacts," Water Resour. Manag., vol. 39, pp. 6701–6714, 2025. doi: 10.1007/s11269-025-04268-6.

[23] W. Y. Oseni, O. A. Akangbe, and K. Abhulimen, "Mathematical modelling and simulation of leak detection system in crude oil pipeline," Heliyon, vol. 9, no. 4, Art. no. e15412, 2023. doi: 10.1016/j.heliyon.2023.e15412.

[24] B. Manshoor, A. Khalid, I. Zaman, D. H. Didane, and N. F. Zulkefli, "Simulation of flow characteristics for leak detection in oil and gas pipeline network," AIP Conf. Proc., vol. 2401, no. 1, p. 020009, 2021. doi: 10.1063/5.0070435.

[25] M. Aghashahi, L. Sela, and M. K. Banks, "Benchmarking dataset for leak detection and localization in water distribution systems," Data in Brief, vol. 48, p. 109148, 2023. doi: 10.1016/j.dib.2023.109148.

[26] D. P. Tavares, M. A. Vasconcellos, S. R. de Farias Neto, and A. G. de Lima, "Evaluation of multiphase flow in the presence of leak in oil pipelines: Modeling and simulation," Defect Diffus. Forum, vol. 348, pp. 64–70, 2014. doi: 10.4028/www.scientific.net/ddf.348.64.

[27] A. Kaveh and N. Khavaninzadeh, "Efficient training of two ANNs using four meta-heuristic algorithms for predicting the FRP strength," Structures, vol. 52, pp. 256–272, 2023. doi: 10.1016/j.istruc.2023.03.178.

[28] K. Bhattacharjee and P. Roy, "Comparative benchmarking of machine learning models for pipeline leak prediction using COMSOL simulation data," 2025. [Manuscript submitted for publication].

[29] T. Yu, X. Zhang, I. E. Lima Neto, T. Zhang, Y. Shao, and M. Ye, "Impact of orifice-to-pipe diameter ratio on leakage flow: An experimental study," Water, vol. 11, no. 10, p. 2189, 2019. doi: 10.3390/w11102189.

[30] M. F. Siddique, Z. Ahmad, and J.-M. Kim, "Pipeline leak detection: A comprehensive deep learning model using CWT image analysis and an optimized DBN-GA-LSSVM framework," Sensors, vol. 24, no. 12, p. 4009, 2024. doi: 10.3390/s24124009.

[31] M. F. Siddique, Z. Ahmad, N. Ullah, S. Ullah, and J.-M. Kim, "Real-time pipeline leak detection: A hybrid deep learning approach using acoustic emission signals," Appl. Sci., vol. 15, no. 1, p. 185, 2025. doi: 10.3390/app15010185.

[32] F. Xi, L. Liu, L. Shan, B. Liu, and Y. Qi, "Pipeline leak identification and prediction of urban water supply network system with deep learning ANN," Water, vol. 16, no. 20, p. 2903, 2024. doi: 10.3390/w16202903.

[33] J. Yang, H. Mostaghimi, R. Hugo, and S. S. Park, "Pipeline leak and volume rate detections through AI and vibration analysis," Measurement, vol. 187, p. 110368, 2022. doi: 10.1016/j.measurement.2021.110368.

[34] A. Kaveh, Applications of Artificial Neural Networks and Machine Learning in Civil Engineering, Studies in Computational Intelligence, vol. 1168. Springer, 2024. doi: 10.1007/978-3-031-66051-1.

[35] P. Sukarno, K. Sidarto, A. Trisnobudi, D. Setyoadi, N. Rohani, and D. Darmadi, "Leak detection modelling and simulation for oil pipeline with artificial intelligence method," ITB J. Eng. Sci., vol. 39B, pp. 1–19, 2007. doi: 10.5614/itbj.eng.sci.2007.39.1.1.

[36] S.-H. Kim, "Multiple leak detection algorithm for pipe network," Mech. Syst. Signal Process., vol. 139, Art. no. 106645, 2020. doi: 10.1016/j.ymssp.2020.106645.

[37] X. Wang and M. S. Ghidaoui, "Identification of multiple leaks in pipeline: Linearized model, maximum likelihood, and super-resolution localization," Mech. Syst. Signal Process., vol. 107, pp. 529–548, 2018. doi: 10.1016/j.ymssp.2018.01.042.

[38] X. Wang and M. S. Ghidaoui, "Identification of multiple leaks in pipeline II: Iterative beamforming and leak number estimation," Mech. Syst. Signal Process., vol. 119, pp. 346–363, 2019. doi: 10.1016/j.ymssp.2018.10.005.

[39] M. Marino, A. Gomes, S. Del Giudice, and G. Giunta, "Integration of multiple leak detection technologies into a single advanced leak detection system," in Proc. 18th Pipeline Technology Conf., Berlin, Germany, 2023.

[40] A. Kaveh, A. D. Eslamlou, S. M. Javadi, and N. G. Malek, "Machine learning regression approaches for predicting the ultimate buckling load of variable-stiffness composite cylinders," Acta Mechanica, vol. 232, no. 3, pp. 921–931, 2021. doi: 10.1007/s00707-020-02878-2.