A Machine-Learning Based Approach to Leak Detection in Water Distribution Networks

Jonathan Weekes and Patrick Hosein
The University of the West Indies
St. Augustine, Trinidad and Tobago
jonathan.weekes1@my.uwi.edu, patrick.hosein@uwi.edu

***Abstract*—Pipeline leaks in water distribution networks cause major losses and infrastructure damage. While traditional detection methods are effective at locating leaks, they remain costly and unsuitable for real-time monitoring. This study develops a machine learning framework for leak detection and localization using pressure data enhanced by signal and statistical feature engineering. Signal transformations (Fourier, Wavelet, Cepstral), statistical metrics, and three classifiers (XGBoost, SVM, and ANN) were evaluated on a benchmark dataset under varied flow conditions. Post optimization, the ANN achieved 97% accuracy under low-flow scenarios, with XGBoost reaching 93% and demonstrating superior robustness. SVM matched peak performance but lacked consistency. A meta-analysis showed this study's ANN, using only pressure data, outperformed a comparable high-flow condition pressure-only model (77% vs. 71.8% accuracy). However, the comparative model using multiple sensor types achieved higher accuracy (86.5%). Nevertheless, the XGBoost model of this research surpassed all benchmarks (89%) due to the richer set of features. Feature importance analysis highlighted the discriminative role of dominant frequencies, wavelet coefficients, and statistical descriptors. Overall, the findings confirm that integrating signal analysis with statistical profiling enables scalable and accurate leak detection in smart water networks.**

***Index Terms*—Leak Detection, Machine Learning, Signal Analysis, Feature Engineering, Smart Water Infrastructure**

I. Introduction

Leaks are the enemy of water supply systems all over the world. According to recent studies, at least 20% of treated freshwater is lost during transportation due to leakages. Beyond the sheer wastage of this indispensable resource, leaks exacerbate environmental and socio-economic vulnerabilities. Service disruptions caused by leaks can severely impact fragile ecosystems, especially in regions already facing water scarcity — a condition that is disproportionately experienced in agriculturally dependent and climate-stressed communities [1]. In the Caribbean country of Trinidad and Tobago for example, the average daily per capita water consumption stands at 82 gallons, nearly double the regional average of 46 gallons per day according to the Chief Executive Officer (Ag.) of the Water and Sewerage Authority (WASA) in 2021. As the population grows, so too will this demand, placing additional strain on aging infrastructure [2]. Traditional leak detection methods such as thermal imaging, acoustic sensors, and ground-penetrating radar (GPR) have demonstrated effectiveness in identifying leaks at localized points within a pipeline network. However, these techniques are often resource-intensive, requiring expensive equipment and highly skilled personnel to operate [3]. In contrast, continuous or static systems rely on sensors and data collectors that are fixed within the water network capable of monitoring various characteristics of the water distribution pipeline in real-time such as pressures and flows. Sensors commonly deployed include accoustic sensors, flow rate monitors and pressure sensors all of which are capable of detecting transient changes in water pipeline systems signaling potential leaks. For optimal performance, static leak detection methods are typically combined with computational pipeline monitoring (CPM) which employs algorithms and models to analyze pipeline data and detect leaks in real time. This study builds on this body of work by investigating the role of signal processing and statistical feature engineering in enhancing ML model performance for CPM. This study addresses three primary objectives: (i) evaluating the effectiveness of digital signal analysis methods in extracting discriminative features from pressure sensor data for early leak detection; (ii) examining the contribution of statistical feature extraction in improving the accuracy and robustness of machine learning models; and (iii) assessing the performance of XGBoost, Support Vector Machines (SVM), and Artificial Neural Networks (ANN) when integrated with combined signal-based and statistical feature engineering for leak detection and localization in water distribution networks.

978-1-6654-7808-3/25/$31.00 © 2025 IEEE

This study relies on a laboratory-scale benchmarking dataset [4], which, while valuable for systematic evaluation, does not fully capture the scale, variability, and leak diversity of real-world systems. The exclusive use of pressure data, despite the availability of other sensing modalities, may also constrain signal richness. Nevertheless, while these factors have the propensity to limit transferability, they are consistent with the exploratory scope of the study and provide a clear direction for future extensions.

II. Literature Review

A. Leak Detection

Researchers have developed numerous methods for detecting water pipeline leaks over the years [5] . [6] reflects this by presenting an extensive list of 20 different leak detection methods, which they broadly classified into two categories: static and dynamic. A similar taxonomy was described by [3], who explained that static leak detection systems rely on fixed sensors or data collectors installed within the water distribution network, often transmitting data periodically to management offices for analysis. These systems enable continuous monitoring to identify, localize, and pinpoint leaks. In contrast, dynamic leak detection systems involve devices deployed to suspected leakage areas to conduct on-site investigations.

Despite their distinctions, researchers agree that optimal leak detection often combines both static and dynamic approaches, as each encompasses a range of technologies that collectively enhance accuracy [3]. For instance, acoustic technologies such as noise loggers can function as dynamic tools when moved periodically across locations, or as static devices when left permanently installed within the network. Building on this principle, the present study employed a pair of dynamic pressure sensors configured statically to monitor pressure variations in the experimental water distribution network.

In contrast, researchers in [7] observed leakage detection methods within a broader categorization framework where they classified leak detection techniques into physical and electronic methods. Physical methods were deemed traditional; requiring significant expertise making them labor-intensive, time consuming and costly. They cited electronic methods on the other hand as the use of modern technologies such as sensors, automated data analytics and artificial intelligence (AI) to detect and analyze leaks. Based on their research, they noted a considerable proportion of recent studies targeting the exploration of AI and Machine Learning based leak detection systems, corroborating the rapidly growing usefulness of such systems. According to [7] however, there are still significant challenges to be addressed in the development of machine learning based leak detection technology, such as improving the accuracy and sensitivity of detection sensors, ensuring consistent performance in various environments, and developing efficient data processing and analysis for feature engineering.

This research attempts to supplement further insights to the body of knowledge concerned with the processing and analysis of raw sensory data, in particular, pressure sensor data for improving the leak detecting performance of machine learning models. It also goes one step further by considering the different classes of leaks to be identified whilst also accounting for the presence of varying background flow conditions.

B. Feature Engineering

Feature engineering involves selecting, transforming, and constructing input variables to convert noisy sensor data into meaningful features that enhance accuracy, reduce computational load, and improve interpretability [8], [9]. According to Emerson's engineers [10], six broad categories of signal analysis are commonly used: time-domain, frequency, order, time-frequency, cepstrum, wavelet, and model-based methods. Time-domain analysis summarizes overall signal behavior using statistical measures such as mean, RMS, variance, and energy, often computed as rolling-window features for both time and frequency domains [11]. Frequency analysis via the Fast Fourier Transform (FFT) decomposes signals but is limited with non-stationary data, where time-frequency methods are more effective. Cepstrum and wavelet analysis further extend this scope where cepstrum analysis detects periodicities, while wavelets excel at capturing transients and have proven valuable in pipeline leak detection [12]. Model-based approaches such as AR and ARIMA forecast time-series patterns, though they were unsuitable here due to the dataset's short data collection duration.

To balance complexity with interpretability, this study emphasized relevant statistical and signal processing techniques, supplemented with Principal Component Analysis (PCA) for dimensionality reduction where necessary. As highlighted in [10] and [13], such methods are crucial for transforming raw sensor data into informative inputs that enhance learning performance. Building on this foundation, the present work applies these approaches to evaluate their effectiveness in distinguishing between leak and non-leak conditions, with the aim of improving detection accuracy while minimizing false positives.

C. Model Selection & Feature Importance

Finally, drawing on the comprehensive review by Pérez and Sofía [14], which examined and synthesized the effectiveness of various machine learning models for water pipeline leak detection, this study focused on the most prominent techniques namely: Support Vector Machines (SVMs), XGBoost, and Artificial Neural Networks (ANNs). This selection is further reinforced by the findings of [11], who noted that much of the existing research frames leak detection as a classification task and predominantly employs ANNs, SVMs, and other tree-based methods to achieve high accuracy.

In addition to robust model selection and evaluation this study also considered model interpretability and explainability, recognizing their importance in understanding the role of key features in leak detection models. Recent studies emphasize the value of explainable AI techniques for this purpose. [15] highlights SHAP (SHapley Additive exPlanations) as a key method that quantifies the contribution of each feature to model predictions, offering both local and global interpretability. Rooted in cooperative game theory, SHAP values overcome limitations of traditional importance measures such as linear coefficients, Gini, or permutation importance, which can be biased by scale dependencies. [16] further underscores SHAP's strength in capturing feature interactions, thereby providing a more consistent and reliable framework for understanding model behavior.

Despite these advances, several challenges persist in applying machine learning to leak detection. A key limitation is model generalization as approaches that perform well in controlled settings often degrade under real-world pipeline variations. Real-time constraints also hinder adoption, since deep learning models with high computational costs may be impractical for large-scale systems requiring immediate response [17]. Data scarcity and labeling difficulties further limit the development of robust classifiers, particularly for rare or intermittent leaks.

By identifying statistical and signal analysis features that can be reliably derived from raw pressure data and pairing them with algorithms capable of generalizing across flow conditions, this paper contributes to ongoing efforts aimed at enabling practical machine learning adoption for pipeline leakage detection, with potential for local applications.

III. Methodology

A. Research Design

This study follows an experimental research design, leveraging both quantitative analysis and the results of empirical testing. To achieve effective quantitative analysis, a standard machine learning-based approach was implemented as outlined in Figure 1.

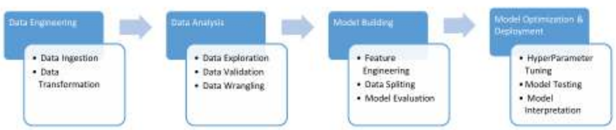

Fig. 1. Basic Data Science and Machine Learning Workflow Implemented

The empirical testing component of this research design was obtained as a dataset of raw but labeled sensor readings made open source by researchers in [4]. Mohsen Aghashahi, Lina Sela, M. and Katherin e Banks, in their paper, *Benchmarking dataset for leak detection and localization in water distribution systems*, published data that were generated via a laboratory-scale water distribution system that included:

1) Three types of sensors: accelerometers, hydrophones, and dynamic pressure sensors
2) Four Leak Types: Orifice leak, Longitudinal Cracks, Circumferential cracks, Gasket leak, and a no-leak condition
3) Two Network Topologies: Looped and Branched
4) Six background flow conditions as characterized by different noise and service demand variations

For the purposes of this research objective, only dynamic pressure sensor data concerned with the branched pipeline topology were utilized. Additionally, only 4 of the 6 background flow conditions were accounted for, as the other two required the use of acoustic data.

B. Data Collection Methods & Experiment Description

The dataset consisted of analog time-series pressure readings recorded for a branched network topology with various operational conditions. These included 4 different background service flow conditions, each measured in liters per second (LPS) as well as 4 artificially induced leak events. Figure 2 illustrates the branched pipeline architecture..

As denoted in Figure 2, the experimental setup simulated background flow using a 25.4 mm service line connected via a saddle clamp representing a demand of 0.44 LPS for approximately 100 people. Daily variability was captured with multipliers producing flows of 0.18 LPS and 0.47 LPS,

Fig. 2. Experimental Water Distribution Test Bed in a Branched Topology Configuration

alongside conditions of zero demand and a transient drop from 0.47 LPS to 0 LPS 20 seconds after the start of the leak experiment. Pressure sensor data were recorded at high sampling rates for approximately 30 seconds per run. To investigate leak characteristics, four leak types: Orifice Leaks (OL), Longitudinal Cracks (LC), Circumferential Cracks (CC), and Gasket Leaks (GL) were induced in separate pipes in the middle of the test bed, while additional experiments with a leak-free pipe provided data for a “no leak” class, ensuring a balanced dataset across varying flow and leak conditions.

C. Analysis Procedure

This study employed a structured methodology to evaluate the effectiveness of statistical and signal analysis-based feature extraction in improving machine learning models for leak detection and localization. Data were first ingested from multiple CSV files, processed into analysis-ready dataframes and preprocessed through label encoding, resampling and smoothing of pressure signals.

A comprehensive feature engineering process was then applied: eleven statistical features were extracted directly from the raw pressure signal namely: pressure range, pressure drop, peak count, RMS, standard deviation, skewness, signal energy, Shannon entropy, signal-to-noise ratio, cross-correlation and autocorrelation. In parallel, three signal analysis methods (FFT, CWT, and Cepstrum) were performed on the time series data, each yielding six complementary statistical descriptors (mean, maximum, minimum, standard deviation, energy and PCA-reduced components) producing a total of 18 signal-analysis derived features.

Feature selection was then conducted using Pearson correlation and variance inflation factor analysis to ensure robustness. The trimmed feature set was then used to train Support Vector Machines (SVM), XGBoost and Artificial Neural Networks (ANN) with performance evaluated through precision, recall, F1-score and accuracy as well as a SHAP-based feature importance analysis. Finally, hyperparameter tuning and stratified K-fold cross-validation were applied to optimize model performance and interpret results.

IV. Results

A. Exploratory Data Analysis

The initial dataset comprised 4,576,576 pressure readings, balanced across the four leak classes investigated. Each reading corresponded to a time increment of 0.1ms from the dynamic pressure sensor yielding highly granular measurements. To enable effective pattern extraction, the data were aggregated to 0.5 second intervals resulting in 302 samples while maintaining comparable class distributions.

Figure 3 shows the density distributions of raw pressure values from the two sensors across different background flow conditions and leak types where *P1* and *P2* refer to the sensors located upstream and downstream of the leak respectively.

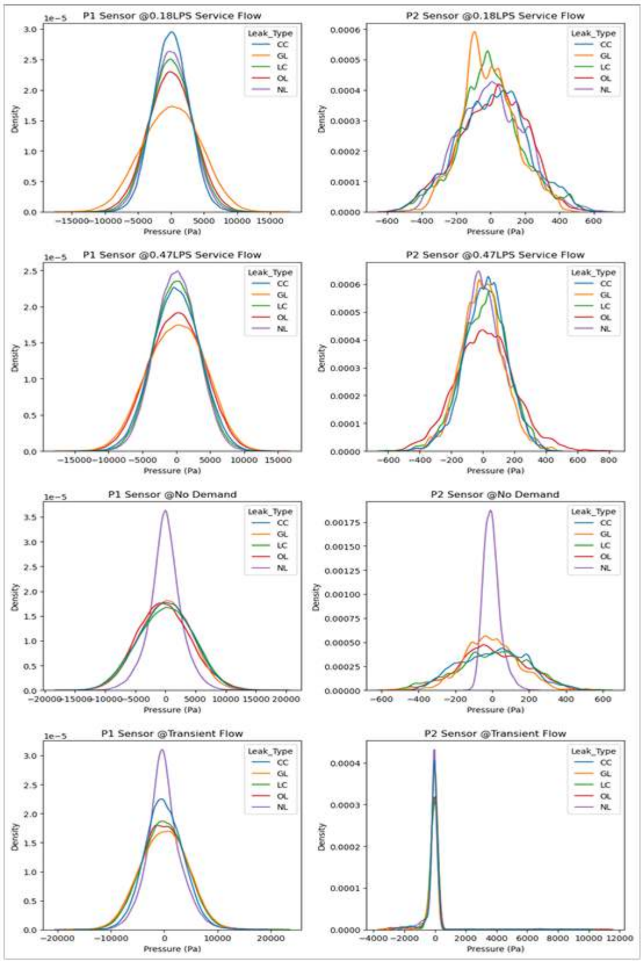

Fig. 3. Density Distribution of Pressure per Leak Type for Different Service Flow Conditions

B. Model Development and Evaluation

Following feature selection, models were developed as outlined in Section 3 and performance was evaluated and compared. Post-optimization results (Table I) show that the ANN achieved the highest overall accuracy across most flow conditions, peaking at 0.97 under 0.18 LPS and outperforming both XGBoost and SVM in transient and no demand scenarios demonstrating strong robustness and generalization. XGBoost was competitive, particularly at lower background flows, achieving the best accuracy (0.89) at 0.47 LPS but its performance declined under transient conditions (0.78).

The SVM model exhibited the weakest performance overall, with notable declines in accuracy under higher flow rates and

TABLE I
PERFORMANCE COMPARISON OF ML MODELS UNDER VARYING FLOW CONDITIONS POST OPTIMIZATION

Flow	Leak	XGBoost				SVM				ANN			
		Prec	Rec	F1	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1	Acc
0.18 LPS	CC	1.00	1.00	1.00		1.00	0.94	1.00		1.00	1.00	1.00	
	GL	1.00	1.00	1.00		0.96	1.00	0.98		0.96	1.00	0.98	
	LC	0.82	0.82	0.82	0.93	0.83	0.88	0.79	0.93	0.94	0.88	0.91	0.97
	NL	1.00	0.89	0.94		0.94	0.89	0.83		0.94	0.94	0.97	
	OL	0.84	0.94	0.89		0.94	0.94	0.94		1.00	1.00	1.00	
0.47 LPS	CC	0.76	0.72	0.74		0.61	0.78	0.68		0.67	0.67	0.67	
	GL	1.00	1.00	1.00		0.84	0.95	0.89		0.84	0.95	0.89	
	LC	0.88	0.82	0.85	0.89	0.54	0.41	0.47	0.74	0.69	0.65	0.67	0.77
	NL	0.80	0.89	0.84		0.81	0.72	0.91		0.63	0.67	0.65	
	OL	1.00	1.00	1.00		0.87	0.76	0.91		0.93	0.76	0.84	
Transient	CC	0.57	0.44	0.50		0.57	0.67	0.62		0.62	0.56	0.59	
	GL	0.95	0.95	0.95		1.00	0.86	0.93		0.95	0.95	0.95	
	LC	0.47	0.53	0.50	0.78	0.53	0.53	0.53	0.80	0.58	0.65	0.61	0.84
	NL	1.00	0.94	0.97		1.00	0.94	0.97		1.00	0.94	0.97	
	OL	0.85	1.00	0.92		0.94	1.00	0.97		0.94	1.00	0.97	
No Demand	CC	0.85	0.94	0.89		0.70	0.89	0.78		0.89	0.94	0.92	
	GL	1.00	0.95	0.98		1.00	0.95	0.98		0.95	0.95	0.95	
	LC	0.69	0.65	0.67	0.86	0.73	0.47	0.57	0.82	0.79	0.65	0.71	0.87
	NL	0.94	0.89	0.91		0.89	0.47	0.89		0.94	0.89	0.91	
	OL	0.78	0.82	0.80		0.76	0.82	0.78		0.75	0.88	0.81	

demand variations. Overall, these findings suggest that while traditional machine learning models like XGBoost and SVM provide reasonable detection capabilities, ANN architectures are better suited for handling the variability inherent in real-world pipeline leak detection.

C. Feature Importance Evaluation

Figure 4 illustrates the relative importance of predictor variables, independent of flow condition, across the three models (XGBoost, SVM, and ANN) as measured by their frequency of selection as significant features. It should be noted that these features emerged as the most suitable features from the initial list of all features engineered. The predictors `P1_Sensor_FFT_maxfreq`, `P1_AutoCorr`, and `P2_Sensor_PeakCount` consistently ranked highest, with XGBoost showing strong reliance on frequency-domain and autocorrelation features. PCA-derived and continuous wavelet transform-based features also contributed moderately, indicating their value in capturing distinctive leak signatures. In contrast, variables such as `P1_P2_PDrop`, `P2_Sensor_Cepstrum_Mean`, and `P1_Sensor_Range` showed limited influence, suggesting minimal impact on classification performance. Overall, signal derived features from frequency and autocorrelation domains dominated across models in discriminating leak conditions.

V. DISCUSSION

A. Proficiency of Digital Signal Analysis Techniques

This research objective evaluated the effectiveness of signal processing transformations — FFT, CWT, cepstral analysis, and autocorrelation—in extracting leak-relevant features from dynamic pressure signals. Across all models

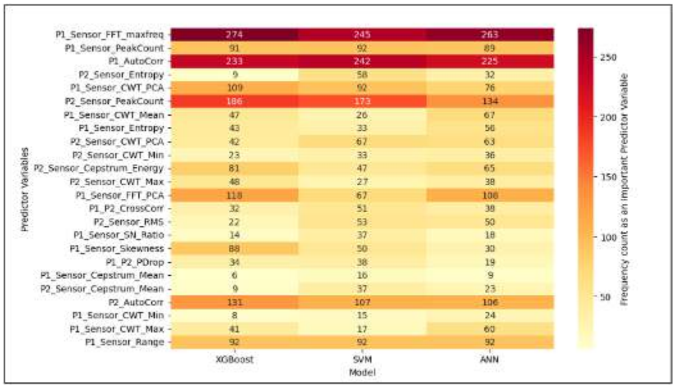

Fig. 4. Feature Importance Heatmap Across All Models based on a SHAP value analysis of Test Predictions

and flow conditions, frequency-domain and autocorrelation features consistently emerged as key predictors, indicating that leaks induce detectable shifts in dominant frequencies and periodic pressure oscillations. Notably, features derived from the Continuous Wavelet Transform (CWT), especially `P1_Sensor_CWT_PCA`, were influential under transient flows, showcasing their ability to capture non-stationary leak signatures through time-localized frequency analysis (Figure 5). Although cepstral features were less robust overall, they provided insights under high-flow conditions by isolating harmonic reflections linked to leak dynamics, albeit necessitating hyperparameter tuning to mitigate noise-related variability. Ultimately, the results confirm that frequency-domain and time-frequency transformations are robust indicators of leak onset, with CWT features offering distinct advantages in transient flow regimes.

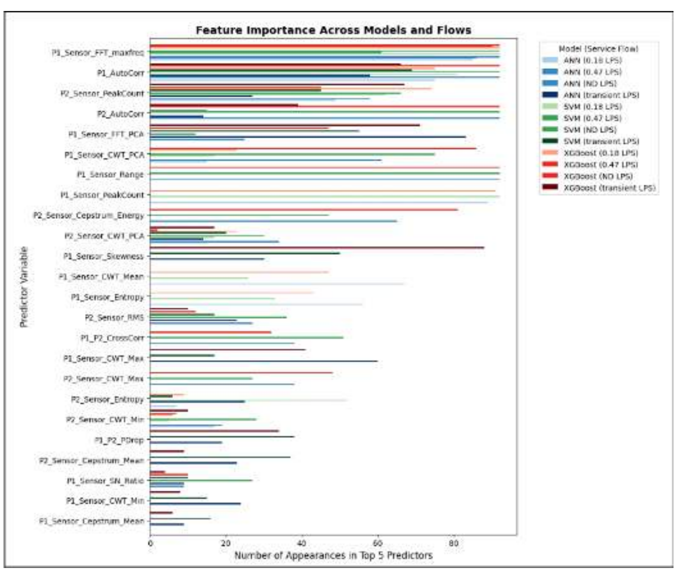

Fig. 5. Feature Importance Across Models and Flows

B. Role of Statistical Feature Extraction

The second research objective highlighted the complementary value of statistical features in improving classification performance. Peak Count consistently emerged as the most influential predictor reflecting its sensitivity to pressure surges induced by leaks. Other features, such as skewness and principle components dimensionality reduction, effectively captured asymmetry and signal complexity proving especially useful under transient flow conditions. Range-based metrics, particularly during off-peak hours, further demonstrated strong predictive power by isolating pressure variations in low-noise environments. Overall, these statistical descriptors enhanced model generalization across diverse operating states and, as confirmed by high SHAP importance rankings, provided distinctive insights beyond those offered by signal transformation techniques alone.

C. Comparative Model Evaluation

The third objective compared the performance of XGBoost, SVM, and ANN for multiclass leak detection under varying flow conditions. Prior to tuning, XGBoost consistently achieved the most reliable results, particularly in low-demand scenarios, owing to its robustness in handling non-linear interactions. The ANN demonstrated strong discriminative ability but was more variable across conditions while SVM showed greater consistency but generally lower accuracy. After hyperparameter tuning, ANN achieved the highest accuracies in most conditions (e.g., 0.97 at 0.18 LPS, 0.87 under no demand), while XGBoost remained the most stable and interpretable, excelling at moderate flows (0.89 at 0.47 LPS). SVM exhibited only modest improvements, with persistent limitations in complex feature spaces. Overall, results confirm the effectiveness of deep learning and ensemble methods when integrated with advanced feature engineering, highlighting ANN's adaptability and XGBoost's reliability as key strengths for scalable leak detection.

D. Meta-Analysis

To better contextualize the findings of this study, a comparative analysis was conducted against the work of [6] who also utilized the *Benchmarking Dataset for Leak Detection and Localization in Water Distribution Systems* made available by [4]. This meta-analysis aims to evaluate the relative efficacy of different feature engineering and model training approaches in achieving accurate leak detection and localization. This study demonstrated that hybrid feature engineering using pressure data alone can achieve competitive performance in leak detection (Figure 6).

This study demonstrated that hybrid feature engineering using pressure data alone can achieve competitive performance in leak detection. The ANN model outperformed a comparable pressure-only model from Mahdi et al. (2024) (77% vs. 71.8%), while XGBoost achieved the highest overall accuracy (89%) and robustness, reflecting its ability to leverage complex signal-based features. However, Mahdi et al.'s multimodal ANN, which fused pressure and vibration data, achieved

Study Component	Current Study (2025)	Mahdi et al. (2024)
Dataset	Aghashahi et al. (2023) benchmark	Aghashahi et al. (2023) benchmark
Signal Smoothing	Yes	No
Resampling	Yes - Resampled to represent data points incrementing by 0.5 seconds	Yes - Resampled to represent data points incrementing by 2.5 milliseconds
Sensor Set	Dynamic Pressure Sensors	Accelerometers + Dynamic Pressure Sensors
Feature Engineering	A combination of Signal + Statistical features (See Figure 4.28)	Auto-correlation coefficient & Signal Energy
Leak Scenarios	Orifice leak (OL), Longitudinal Crack; (LC), Circumferential Crack (CC) and Gasket leak (GL)	Orifice leak (OL), Longitudinal Crack; (LC), Circumferential Crack (CC) and Gasket leak (GL)
Background Flow Conditions	0.18LPS, 0.47LPS, Transient & No Demand	0.47 LPS only

Fig. 6. Differences in experimental design between the current study and [6]

superior performance (up to 86.5%), underscoring the limits of pressure-only approaches. Differences in neural network architecture, train-test splits, and optimization methods also contributed to performance variations between studies.

The findings confirm that combining statistical and signal-based features provides strong separability between leak types, with ANN excelling in adaptability and XGBoost in stability and interpretability. Nonetheless, future improvements will likely depend on multimodal data integration and advanced deep learning architectures tailored for time-series analysis. Expanding evaluations to real-world datasets and diverse network configurations remains essential for ensuring scalability and practical deployment.

VI. Conclusion

Notwithstanding the limitations outlined in Section 1, this study provides several methodological and practical contributions to smart water infrastructure, particularly in advancing machine learning-based leak detection through enhanced feature engineering. A hybrid signal processing approach integrating Fourier, Wavelet, and Cepstral analysis with statistical descriptors enabled richer leak signature extraction from pressure data. Comparative evaluation of XGBoost, Support Vector Machines (SVM), and Artificial Neural Networks (ANN) offered insights into classifier performance under varying leak and flow conditions. Feature selection techniques, including PCA and SHAP, emphasized the value of combining statistical and signal-based attributes. Furthermore, the use of a publicly available benchmark dataset strengthens reproducibility and fosters comparative studies. By contextualizing findings within a meta-analysis of recent work [6], the study provides a valuable reference point for evaluating model architectures, sensor configurations, and feature engineering strategies. Importantly, the reliance solely on pressure data demonstrates a cost-effective and scalable pathway for leak detection in resource-limited contexts.

Overall, the findings confirm that integrating signal dynamics and statistical profiling with machine learning constitutes a viable and scalable approach for intelligent leak detection. This lays the groundwork for future research focused on real-time deployment, exploration of deep learning models, and integration with hydraulic simulations and multi-modal sensing with the potential to enhance predictive accuracy, reduce non-revenue water losses and improve global water distribution efficiency.

References

[1] A. V. Serafeim, N. T. Fourniotis, R. Deidda, G. Kokosalakis, and A. Langousis, “Leakages in water distribution networks: Estimation methods, influential factors, and mitigation strategies—a comprehensive review,” *Water*, vol. 16, no. 11, p. 1534, 2024.

[2] R. Ganaselal, “Water and sewerage authority (wasa) report,” Water and Sewerage Authority (WASA), Trinidad and Tobago, 2021, internal report.

[3] S. El-Zahab and T. Zayed, “Leak detection in water distribution networks: an introductory overview,” *Smart Water*, vol. 4, no. 1, 2019.

[4] 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.

[5] M. N. M. Peng, “Water pipelines failures (effective factors),” https://www.linkedin.com/pulse/inspection-effective-factors-water-pipelines-failures-mahdi-noori-a-/, Jul. 2023.

[6] N. M. Mahdi, A. H. Jassim, S. A. Abulqasim, A. Basem, A. a. F. Ogaili, and L. A. Al Haddad, “Leak detection and localization in water distribution systems using advanced feature analysis and an artificial neural network,” *Desalination and Water Treatment*, vol. 320, p. 100685, 2024.

[7] S. Lee and B. Kim, “Machine learning model for leak detection using water pipeline vibration sensor,” *Sensors*, vol. 23, no. 21, p. 8935, 2023.

[8] MSOE University, “The importance of feature engineering in machine learning,” https://online.msoe.edu/engineering/blog/importance-of-feature-engineering-in-machine-learning, Jun. 2024.

[9] DataCamp, “Feature extraction in machine learning: A complete guide,” https://www.datacamp.com/tutorial/feature-extraction-machine-learning, Feb. 2025.

[10] National Instruments, “Vibration analysis and signal processing in labview,” https://www.ni.com/en/shop/labview/vibration-analysis-and-signal-processing-in-labview.html, Nov. 2024.

[11] W. Wang, H. Sun, J. Guo, L. Lao, S. Wu, and J. Zhang, “Experimental study on water pipeline leak using in-pipe acoustic signal analysis and artificial neural network prediction,” *Measurement*, vol. 186, p. 110094, 2021.

[12] N. V. Rad, H. A. Nadian, R. Ranzi, M. Rahmanshahi, and M. S. Bejestan, “Investigation of pressure signal and leak detection in pipes by using wavelet transform in transient flow,” in *International Joint Conference on Water Distribution Systems Analysis & Computing and Control for the Water Industry*, vol. 72, 2024, p. 76.

[13] DataCamp, “A data scientist’s guide to signal processing,” https://www.datacamp.com/tutorial/a-data-scientists-guide-to-signal-processing, Aug. 2023.

[14] A. P. Pérez and A. S. Sofía, “A review of leak detection and prediction methods in water distribution systems using machine learning,” https://repositorio.uniandes.edu.co/entities/publication/f7625e58-edd7-4ec3-a419-f0bc50456757, Mar. 2023.

[15] H. Dandamudi, “Exploring Explainable AI: A Deep Dive into SHAP,” Medium, Dec. 2024, accessed: [insert access date]. [Online]. Available: https://medium.com/@hareesha1906/exploring-explainable-ai-a-deep-dive-into-shap-eb06be1a4197

[16] S. Yadav, “Explainable AI (XAI): The Future of Transparent Machine Learning,” Medium, Major Digest, Sep. 2024, accessed: [insert access date]. [Online]. Available: https://medium.com/majordigest/explainable-ai-xai-the-future-of-transparent-machine-learning-a0c8a036d367

[17] X. Zhang, J. Shi, M. Yang, X. Huang, A. S. Usmani, G. Chen, J. Fu, J. Huang, and J. Li, “Real-time pipeline leak detection and localization using an attention-based lstm approach,” *Process Safety and Environmental Protection*, vol. 174, pp. 460–472, 2023.