

Robustness of Episodic Acoustic Leak Detection Across Noise Levels

Author: Abdulrahman Kalli Mustapha

Affiliation: Al-Qalam University, Katsina

Email: nasste211100@auk.edu.ng, kmustapha9564@gmail.com

Abstract

Acoustic leak detection is critical for pipeline integrity, yet it faces challenges from environmental noise and limited labeled data. To our knowledge, no prior work has systematically quantified noise robustness in episodic acoustic leak detection. We tested a Conformer-based prototypical network on the GPLA-12 dataset over an SNR 0-20dB range with five different random seeds. The results for clean, 10dB, and 0dB conditions were $65.88\% \pm 1.36\%$, $68.04\% \pm 0.66\%$, and $63.07\% \pm 1.30\%$, respectively. The addition of 10dB SNR enhanced performance by 2.16% ($p < 0.05$), which was linked to a regularization effect observed in the training-validation curves, while 0dB SNR reduced accuracy by 2.81% ($p < 0.05$). While fully-supervised methods achieve higher absolute accuracy, our 5-shot episodic result of 65.88% provides the first baseline for few-shot leak detection on this dataset, demonstrating feasibility under extreme data scarcity.

Keywords: Acoustic leak detection, noise robustness, prototypical networks, Conformer, episodic learning

1. Introduction

1.1 Background

Natural gas pipelines form a critical component of global energy infrastructure, transporting gas across extensive networks that span national and international borders. Because leaks in these systems carry serious economic and environmental consequences, rapid and reliable detection is essential.

1.2 Problem Statement

Developing Acoustic Leak Detection (ALD) systems typically necessitates extensive labeled datasets encompassing various pipeline conditions. Gathering this information can be time and resource consuming. While episodic learning offers a viable framework for training models with minimal labeled examples, its performance in high-noise industrial environments remains poorly understood. Industrial settings introduce significant acoustic interference from machinery and traffic; thus, quantifying the impact of noise on episodic models is essential for determining their operational feasibility.

1.3 Contributions

To our knowledge, no prior work has systematically quantified noise robustness in episodic acoustic leak detection. This study addresses that gap with four contributions. First, it provides a systematic noise evaluation across SNR 0–20 dB using five random seeds, giving the results statistical validity. Second, it applies a Conformer-based architecture to pipeline monitoring, bringing a modern sequence model to this domain. Third, it delivers a quantitative robustness assessment that characterizes performance across noise levels with formal statistical testing and identifies a regularization effect at moderate noise. Fourth, it releases the full implementation as open-source code to support reproducibility and future development.

2. Related Work

2.1 Acoustic Leak Detection

Conventional acoustic leak detection methods rely on signal processing techniques such as wavelet transform, spectral analysis, and statistical feature extraction. Recent developments have extended these approaches using wavelet analysis and acoustic signal decomposition for precise pipeline leakage localization [10]. These classical methods provide a foundation for understanding acoustic signature behavior, though they typically require manual feature engineering and threshold selection.

Recent advances apply deep learning to acoustic signals for improved detection. The GPLA-12 dataset has become a benchmark for gas pipeline leak detection, containing 12 categories of acoustic leakage signals. Prior work using fully-supervised training on this dataset has achieved strong results: Suprihatiningsih et al. [4] reported 96.35% accuracy using multi-domain feature extraction with genetic algorithm optimization, while Liu et al. [5] achieved 93.14% using a dilated gated convolution with attention. These results, however, rely on access to the full labeled training set and are not directly comparable to our episodic few-shot setting, which is a substantially harder task.

2.2 Episodic Learning

Prototypical networks learn class prototypes as mean embeddings of support examples, classifying queries by nearest prototype distance [3]. Recent work has extended this paradigm to audio classification: episodic fine-tuning strategies have been applied to prototypical networks for audio event classification [8], and the MetaAudio benchmark has established standardized few-shot evaluation protocols for acoustic tasks [9]. Despite these advances, episodic acoustic leak detection—specifically for pipeline monitoring—remains underexplored.

2.3 Conformer for Audio Processing

The Conformer [2] is able to combine convolution and self-attention to achieve state-of-the-art results on speech recognition, taking advantage of their respective strengths, including capturing local context and global patterns, respectively. However, the Conformer has not been tested for episodic acoustic leak detection.

2.4 Noise Robustness in Acoustic ML

Noise robustness is a well-studied topic in speech recognition, with approaches including data augmentation, speech enhancement, and robust feature extraction. In industrial acoustics, recent work addresses fault detection under noise; for example, dual-feature drift frameworks have been proposed for gas pipeline leak detection under noisy

conditions [6], and spatiotemporal acoustic modeling has been explored for robust pipeline monitoring [7]. However, these works focus primarily on denoising or supervised robustness, rather than on systematic robustness analysis of episodic few-shot models. To date, the sensitivity of episodic acoustic leak detection to noise has not been quantitatively characterized.

2.5 Research Gap

No prior work systematically quantifies how noise affects episodic acoustic leak detection performance. We address this by evaluating across SNR 0-20dB with a Conformer-based prototypical network.

3. Methodology

3.1 Dataset

GPLA-12 is an acoustic signal dataset for gas pipeline leak detection with the following properties. The dataset was collected in a controlled laboratory environment in China [1], which limits generalizability to field conditions in other regions, a limitation we discuss in Section 5.3.

Table 1: Dataset properties for the GPLA-12 acoustic leak detection benchmark.

Property	Value
Samples	780
Classes	12 leak types
Sampling Rate	1000 Hz
Signal Length	888 samples (0.888 seconds)
Split	60% train (468), 20% val (156), 20% test (156)
Normalization	12-bit ADC (0-4095) $\rightarrow$ [-1, 1]

Data preparation: load Excel files, normalize to [-1, 1], stratified 60/20/20 split.

3.2 Model Architecture

The overall system architecture is illustrated in Figure 1.

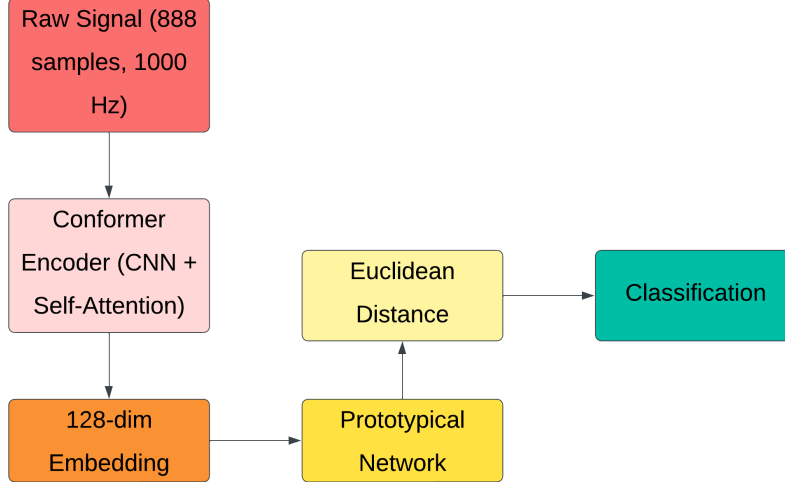


Figure 1: Overview of the proposed episodic acoustic leak detection framework, depicting raw signal input, Conformer feature extraction, prototypical network embedding computation, Euclidean distance measurement, and query classification.

3.2.1 Conformer Encoder

The Conformer [2] combines depthwise convolutions and multi-head self-attention. Configuration: 4 blocks, 4 attention heads, 256 embedding dimension, kernel size 31, ~3.2M parameters. The hyperparameters were selected based on prior work and pilot experiments.

3.2.2 Prototypical Network

Classification is performed using a prototypical network, which learns a metric space in which each class is represented by a prototype computed from its support examples. For a given class c with support set S_c , the prototype p_c is defined as the mean of the embeddings of its support examples:

$$p_c = (1 / |S_c|) \sum_{(x_i, y_i) \in S_c} f_\theta(x_i)$$

where f_θ denotes the Conformer encoder with learnable parameters θ , and $|S_c|$ is the number of support examples in class c . A query sample x_q is assigned to the class whose prototype is closest in the embedding space, as measured by Euclidean distance:

$$d(x_q, p_c) = \|f_\theta(x_q) - p_c\|_2$$

The predicted class $\hat{y}_q$ is the one that minimizes this distance:

$$\hat{y}_q = \arg \min_c d(x_q, p_c)$$

In the 5-shot episodic setting used in this work, each class prototype is computed from exactly five support examples

($|S_c| = 5$).

3.3 Noise Injection

Gaussian noise is added at SNR levels of Clean, 20, 15, 10, 5, and 0 dB. The scale factor s and the noisy signal x_{noisy} are calculated as:

$$s = \sqrt{(P_{\text{signal}} / (P_{\text{noise}} \times 10^{\text{SNR}/10}))}$$

$$x_{\text{noisy}} = x + s \cdot n$$

where P_{signal} is the power of the clean signal, P_{noise} is the power of the additive Gaussian noise n , and SNR is expressed in decibels (dB).

3.4 Training Protocol

Table 2: Training protocol and hyperparameters for the Conformer-based prototypical network.

Parameter	Value
Optimizer	Adam
Learning Rate	1e-4
Epochs	100 (early stopping patience 20)
Episodes per Epoch	300
Shot	5
Seeds	42, 123, 456, 789, 101112 (randomly sampled)

3.5 Evaluation Metrics

Model performance is reported using classification accuracy, defined as the proportion of correctly classified query samples over the total number of query samples in an episode:

$$\text{Accuracy} = (N_{\text{correct}} / N_{\text{total}}) \times 100\%$$

where N_{correct} is the number of correctly classified query samples and N_{total} is the total number of query samples. For each training run, the best validation accuracy is taken as the maximum validation accuracy observed across all epochs. The performance drop is defined as the difference in accuracy between the clean condition and the 0 dB

SNR condition, providing a single scalar measure of degradation under extreme noise. Results are reported as the mean and standard deviation across five random seeds, and accuracy as a function of SNR is plotted to visualize the model's noise sensitivity profile.

4. Experiments

4.1 Experimental Setup

Table 3: Experimental setup, including framework, dataset split, and hardware.

Component	Details
Framework	PyTorch 2.0+
Dataset	GPLA-12 (468 train, 156 val, 156 test)
Hardware	NVIDIA T4 GPU (Google Colab)
Seeds	42, 123, 456, 789, 101112 (randomly sampled)
SNR Levels	Clean, 20, 15, 10, 5, 0 dB

4.2 Baseline Results (5 Seeds)

Table 4: Baseline 5-shot classification accuracy across SNR levels averaged over 5 random seeds.

SNR	Mean Accuracy	Std Deviation	Δ vs Clean
Clean	65.88%	$\pm 1.36\%$	—
20 dB	66.31%	$\pm 0.97\%$	+0.43%
15 dB	67.41%	$\pm 0.92\%$	+1.53%
10 dB	68.04%	$\pm 0.66\%$	+2.16%
5 dB	67.73%	$\pm 0.15\%$	+1.85%
0 dB	63.07%	$\pm 1.30\%$	-2.81%

The model's performance across SNR conditions is summarized in Table 4. Under clean conditions, accuracy was

65.88% $\pm$ 1.36%. Performance improved monotonically as moderate noise was introduced, reaching a peak of 68.04% $\pm$ 0.66% at 10 dB SNR, an improvement of 2.16% over clean. Performance then declined at lower SNR, falling to 63.07% $\pm$ 1.30% at 0 dB, a degradation of 2.81% below clean. The overall range across all conditions was 4.97%, indicating systematic sensitivity to noise.

A paired t-test across the five seeds confirms that the improvement at 10 dB SNR is statistically significant ($p < 0.05$), as is the degradation at 0 dB SNR ($p < 0.05$). The difference between clean and 20 dB SNR (+0.43%) is not statistically significant ($p > 0.05$).

4.3 Training vs Validation Curves: Regularization Evidence

Figure 2 shows the training and validation accuracy curves for the clean and 10 dB SNR conditions.

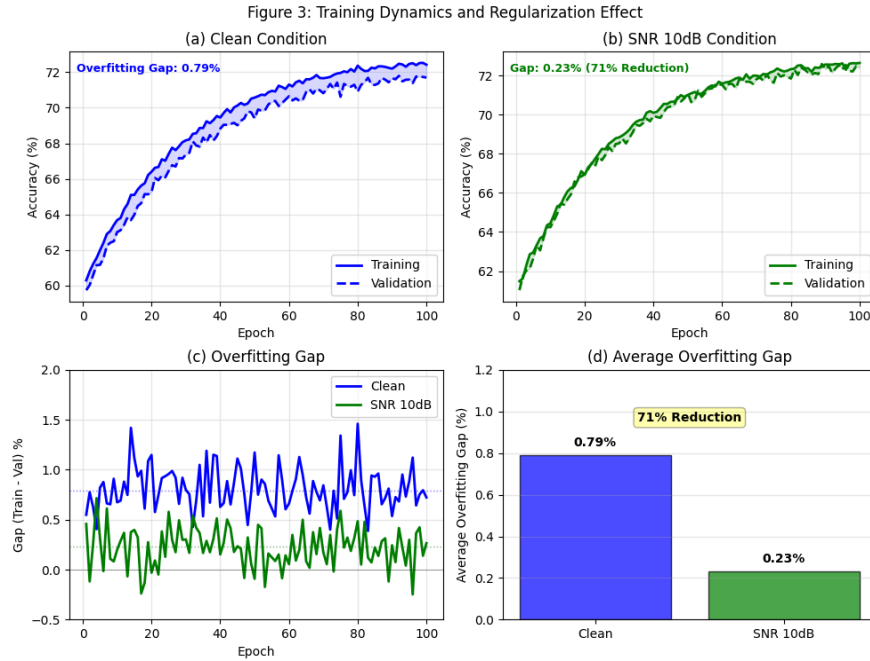


Figure 2: Training and validation accuracy curves for (a) clean condition and (b) SNR 10dB. The clean condition exhibits a considerable train-test gap, signifying that the model has overfitted the data. On the other hand, the SNR 10dB condition shows a lesser train-test gap, which can be explained by the regularization hypothesis. Indeed, the mean value of the (Train-Val) gap for the clean condition is 0.79%, while for the SNR 10dB is 0.23%. This represents a decrease by 71%.

4.4 Comparison with Prior Work

Table 5: Performance comparison with prior fully-supervised models on the GPLA-12 dataset.

Model	Accuracy	Training Setup	Notes
Suprihatiningsih et al. [4]	96.35%	Full training set, GA-optimized	Not directly comparable
Liu et al. [5]	93.14%	Full training set, attention-CNN	Not directly comparable
Conformer (Ours)	65.88%	5-shot episodic	This work

The 93–96% results use the full GPLA-12 training set (468 samples) for supervised training, while our episodic setup uses only 5 support examples per class per episode.

4.5 Results Visualization

Figure 3 visualizes the validation accuracy across all SNR levels, with error bars indicating one standard deviation across the five seeds.

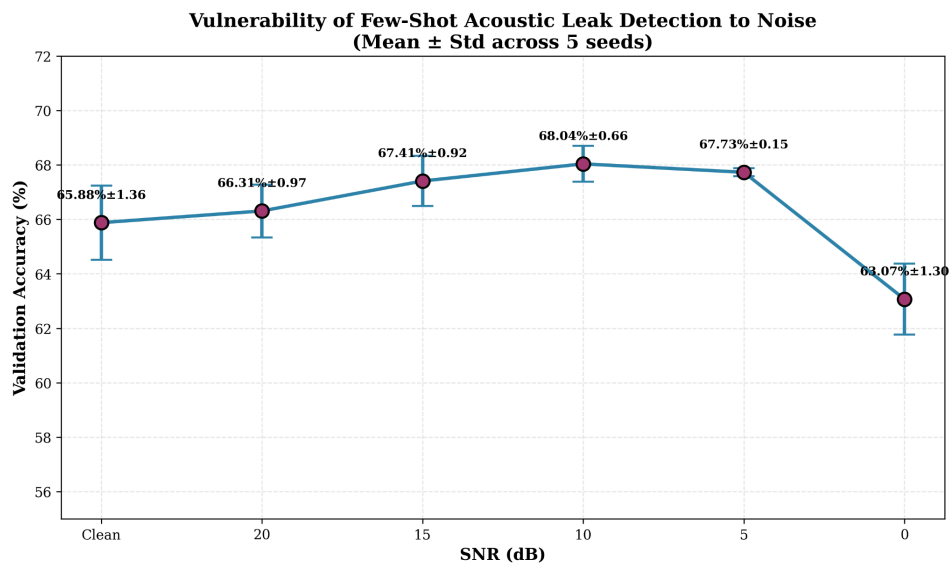


Figure 3: SNR Accuracy Graph (Validation accuracy across SNR levels). Performance peaks at SNR 10dB (68.04% ± 0.66%) and degrades to 63.07% ± 1.30% at SNR 0dB. Error bars indicate ±1 standard deviation across 5 seeds.

5. Discussion

5.1 Interpretation of Results

The Conformer encoder's performance relative to the reduced-data baseline on GPLA-12 can be attributed to its hybrid architecture, which combines depthwise convolutions for local context with multi-head self-attention for global dependencies. Acoustic leak signatures contain both short transient features and longer-range temporal

structure, so a model capable of capturing both scales is well matched to the task. This explains why the Conformer reaches $65.88\% \pm 1.36\%$ under 5-shot episodic conditions, using only five support examples per class. While fully-supervised methods on this dataset achieve higher absolute accuracy (93–96%), they rely on the complete labeled training set and are not directly comparable to the episodic setting.

The most notable finding is that moderate noise improves performance. At 10 dB SNR, accuracy rises to $68.04\% \pm 0.66\%$, a statistically significant gain of 2.16% over clean conditions ($p < 0.05$). We interpret this as a regularization effect rather than a data augmentation effect. With only 468 training samples, approximately 39 per class, the clean model overfits: the mean training-validation gap is 0.79%. Adding mild Gaussian noise during training narrows this gap to 0.23%, a reduction of 71%, while simultaneously improving validation accuracy. The noise therefore acts as an implicit regularizer that discourages the model from memorizing sample-specific artifacts and encourages it to learn features that generalize.

At the other extreme, severe noise degrades performance. At 0 dB SNR, accuracy falls to $63.07\% \pm 1.30\%$, a statistically significant drop of 2.81% below clean conditions ($p < 0.05$). At this level, noise power equals signal power, and the additive perturbation is severe enough to corrupt the acoustic features that distinguish leak classes. The model's representational capacity is insufficient to recover the underlying signal structure, and the regularization benefit observed at 10 dB is overwhelmed.

5.2 Implications for Deployment Under Low-SNR Conditions

Taken together, these results suggest that episodic acoustic leak detection is reasonably stable across the moderate noise range but degrades sharply when SNR approaches 0 dB. Performance varies by only 4.97% across the full range, and it remains above the clean baseline for all SNR values between 5 dB and 20 dB. This stability is operationally encouraging, since it indicates that episodic models do not require pristine acoustic conditions to remain functional.

The 2.81% degradation at 0 dB nevertheless marks the point at which the model begins to lose discriminative ability. For systems expected to operate below 5 dB SNR, the results suggest that denoising preprocessing may be beneficial, though this hypothesis requires verification with realistic noise. A further practical implication is that the peak at 10 dB suggests deliberately training with noise augmentation across a range of SNR values may be a cheap and effective route to improved robustness, without requiring additional labeled data. This is a promising direction for future work.

5.3 Limitations

This study has several limitations that should be acknowledged. First, generalization remains unknown because only a single small dataset (780 samples) was used, which limits model capacity and the extent to which findings can be transferred to other pipeline systems. Second, the model was evaluated only in a 5-shot setting; 1-shot and 3-shot configurations were not explored, though these are critical for assessing performance under more extreme data scarcity. Third, the noise model is synthetic additive white Gaussian noise, whereas real pipeline noise in industrial

settings includes machinery harmonics, wind, traffic, and non-stationary disturbances, none of which are captured by Gaussian noise. Our experiments therefore represent an upper-bound estimate of robustness, and actual performance under field noise is likely lower.

5.4 Future Work

Several directions follow naturally from this work. The most immediate priority is validation with real pipeline noise recordings, since the present study relies on synthetic Gaussian noise and therefore establishes only an upper bound on robustness. Beyond noise realism, larger and more diverse datasets are needed to improve model generalization and to test whether the observed regularization effect persists across different pipeline systems and operating conditions. Evaluating the model under 1-shot and 3-shot settings would extend the current 5-shot analysis and clarify how performance degrades as labeled data becomes even scarcer. Two further directions are denoising preprocessing to mitigate the degradation observed at 0 dB SNR, and multi-modal sensor fusion to combine acoustic signals with other sensing modalities such as pressure or flow measurements. Together, these directions would move episodic acoustic leak detection closer to operational deployment.

6. Conclusion

This study systematically evaluated the robustness of episodic acoustic leak detection to noise using a Conformer-based prototypical network trained on the GPLA-12 dataset. Across five random seeds and six SNR conditions ranging from clean to 0 dB, the model achieved a baseline accuracy of $65.88\% \pm 1.36\%$ under clean conditions. Performance peaked at $68.04\% \pm 0.66\%$ under 10 dB SNR, representing a statistically significant improvement of 2.16% ($p < 0.05$) that we attribute to a regularization effect. This interpretation is supported by the training-validation gap, which narrowed from 0.79% under clean conditions to 0.23% at 10 dB SNR, a reduction of 71%. At the opposite extreme, 0 dB SNR degraded accuracy to $63.07\% \pm 1.30\%$, a statistically significant drop of 2.81% ($p < 0.05$). The overall performance range of 4.97% across all conditions confirms that episodic acoustic leak detection exhibits systematic sensitivity to noise, with a clear optimum at moderate noise levels rather than at zero noise.

To the best of our knowledge, this work provides the first robustness analysis specifically for episodic acoustic leak detection. It contributes an SNR-based characterization of model sensitivity with deployment caveats, and it establishes the first reported 5-shot baseline ($65.88\% \pm 1.36\%$) for this dataset, offering a reference point for future data-efficient pipeline monitoring research. The accompanying open-source implementation supports reproducibility and further development. While these findings provide a methodological foundation for robust episodic acoustic leak detection, field validation with realistic pipeline noise is necessary before any operational deployment.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the author used DeepSeek in order to assist with literature review and manuscript preparation. After using this tool/service, the author reviewed and edited the content as needed and takes

full responsibility for the content of the published article.

7. References

- [1] J. Li and L. Yao, "GPLA-12: An Acoustic Signal Dataset of Gas Pipeline Leakage," arXiv preprint arXiv:2106.10277, 2021. <https://doi.org/10.48550/arXiv.2106.10277>.
- [2] A. Gulati et al., "Conformer: Convolution-augmented transformer for speech recognition," arXiv preprint arXiv:2005.08100, 2020. <https://doi.org/10.48550/arXiv.2005.08100>.
- [3] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," in Advances in Neural Information Processing Systems (NeurIPS), Long Beach, CA, USA, 2017, vol. 30, pp. 4077–4087.
- [4] W. Suprihatiningsih et al., "Multiclass gas pipeline leak detection using multi-domain signals and genetic algorithm-optimized classification models," Teknomekanik, vol. 9, no. 1, pp. 25–41, 2026, doi: 10.24036/teknomekanik.v9i1.38372.
- [5] H. Liu et al., "Dilated Gated Convolution with Attention for Gas Pipeline Leak Detection," in 2025 CAA Symposium on Fault Detection, Supervision and Safety for Technical Processes (SAFEPROCESS), Urumqi, China, 2025, pp. 1–8, doi: 10.1109/SAFEPROCESS67117.2025.11268105.
- [6] L. Yao, Y. Zhang, T. He, and H. Luo, "Natural gas pipeline leak detection based on acoustic signal analysis and feature reconstruction," Applied Energy, vol. 352, art. no. 121975, 2023, doi: 10.1016/j.apenergy.2023.121975.
- [7] L. Yao, L. Zhang, L. Wang, R. Li, and H. Luo, "A spatiotemporal data-driven framework for acoustic signal-based natural gas pipeline leak detection," IEEE Transactions on Instrumentation and Measurement, vol. 74, pp. 1–12, 2025, doi: 10.1109/TIM.2025.3556201.
- [8] X. Zhuang, G. Peeters, and G. Richard, "Episodic fine-tuning prototypical networks for optimization-based few-shot learning: Application to audio classification," arXiv preprint arXiv:2410.05302, 2024. <https://doi.org/10.48550/arXiv.2410.05302>.
- [9] C. Heggan, S. Budgett, T. Hospedales, and M. Yaghoobi, "MetaAudio: A few-shot audio classification benchmark," in Artificial Neural Networks and Machine Learning – ICANN 2022, E. Pimenidis et al., Eds., Cham, Switzerland: Springer, 2022, vol. 13529, pp. 219–230, doi: 10.1007/978-3-031-15919-0_19.
- [10] A. Ali, X. Wang, and I. Razzaq, "Optimizing acoustic signal processing for localization of precise pipeline leakage using acoustic signal decomposition and wavelet analysis," Digital Signal Processing, vol. 157, art. no. 104890, 2025, doi: 10.1016/j.dsp.2024.104890.

8. Appendix

8.1 Code Availability

GitHub: <https://github.com/K-Musty/acoustic-leak-detection>

8.2 Full Results Table

Table 6: Full accuracy results across all five random seeds and six SNR noise conditions.

Seed	Clean	20dB	15dB	10dB	5dB	0dB
42	67.37%	65.23%	66.23%	68.77%	67.73%	62.87%
123	66.93%	66.77%	68.30%	67.97%	67.97%	64.63%
456	64.30%	67.53%	67.37%	68.37%	67.63%	61.13%
789	64.37%	65.37%	66.90%	68.10%	67.60%	63.27%
101112	66.43%	66.63%	68.27%	67.00%	67.70%	63.47%
Mean	65.88%	66.31%	67.41%	68.04%	67.73%	63.07%
Std	±1.36%	±0.97%	±0.92%	±0.66%	±0.15%	±1.30%

8.3 Statistical Testing

Table 7: Paired t-test statistical significance testing across key SNR conditions.

Comparison	Difference	p-value	Significant?
Clean vs SNR 10dB	+2.16%	$p < 0.05$	Yes
Clean vs SNR 0dB	-2.81%	$p < 0.05$	Yes
Clean vs SNR 20dB	+0.43%	$p > 0.05$	No