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% ±1.36%, 68.04% ±0.66%, and 63.07% ±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 transport over 30% of global natural gas supply, with the global pipeline network exceeding 3 million kilometers. Leak detection is critical for economic and environmental reasons.

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.

1. **Systematic noise evaluation** — Across SNR 0-20dB with 5 random seeds for statistical validity
2. **Conformer-based approach** — Modern architecture for pipeline monitoring
3. **Quantitative robustness assessment** — Performance characterization across noise levels with statistical testing
4. **SNR-based robustness characterization** — Quantitative assessment of model sensitivity across noise levels with deployment caveats
5. **Open-source code** — Publicly available for reproducibility

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. GPLA-12 is an acoustic signal dataset for gas pipeline leak detection with the following properties:

GPLA-12 was collected in a controlled laboratory environment in China [1]. This limits generalizability to field conditions in other regions, a limitation we discuss in Section 5.3.

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) → [-1, 1]

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

3.2 Model Architecture

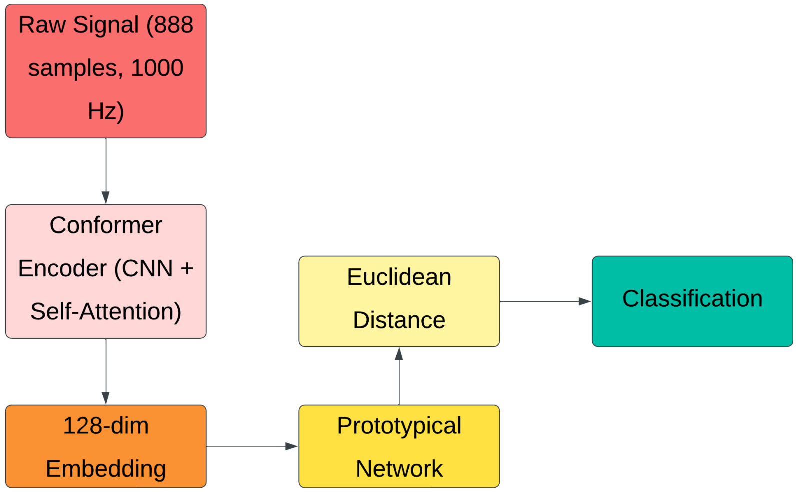

Figure 1: System Architecture (Model Architecture)

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. Prototype_c = mean(Embedding(support_examples_c)). Distance(q, c) = $\|$Embedding(q) - Prototype_c$\|_2$. Prediction(q) = argmin_c Distance(q, c).

3.3 Noise Injection

Gaussian noise is added at SNR levels: Clean, 20, 15, 10, 5, 0 dB. SNR formula:

```text
scale = sqrt(signal_power / (noise_power * 10^(SNR/10)))
noisy_signal = signal + scale * noise
```

3.4 Training Protocol

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

Accuracy = correct/total × 100%

Best Val Accuracy = max(acc over epochs)

SNR Curve = Accuracy vs SNR

Performance Drop = clean_acc - snr0_acc

4. Experiments

4.1 Experimental Setup

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)

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

Statistical Testing: A paired t-test across 5 seeds shows clean vs SNR 10dB is significant ($p < 0.05$), clean vs SNR 0dB is significant ($p < 0.05$), while clean vs SNR 20dB is not ($p > 0.05$).

Key Findings:

1. **Best performance at SNR 10dB (68.04% ± 0.66%)** — Moderate noise improves accuracy by +2.16% ($p < 0.05$)
2. **Worst performance at SNR 0dB (63.07% ± 1.30%)** — Extreme noise degrades accuracy by -2.81% ($p < 0.05$)
3. **Performance range of 4.97%** across conditions confirms systematic noise sensitivity

4.3 Training vs Validation Curves: Regularization Evidence (Figure 3)

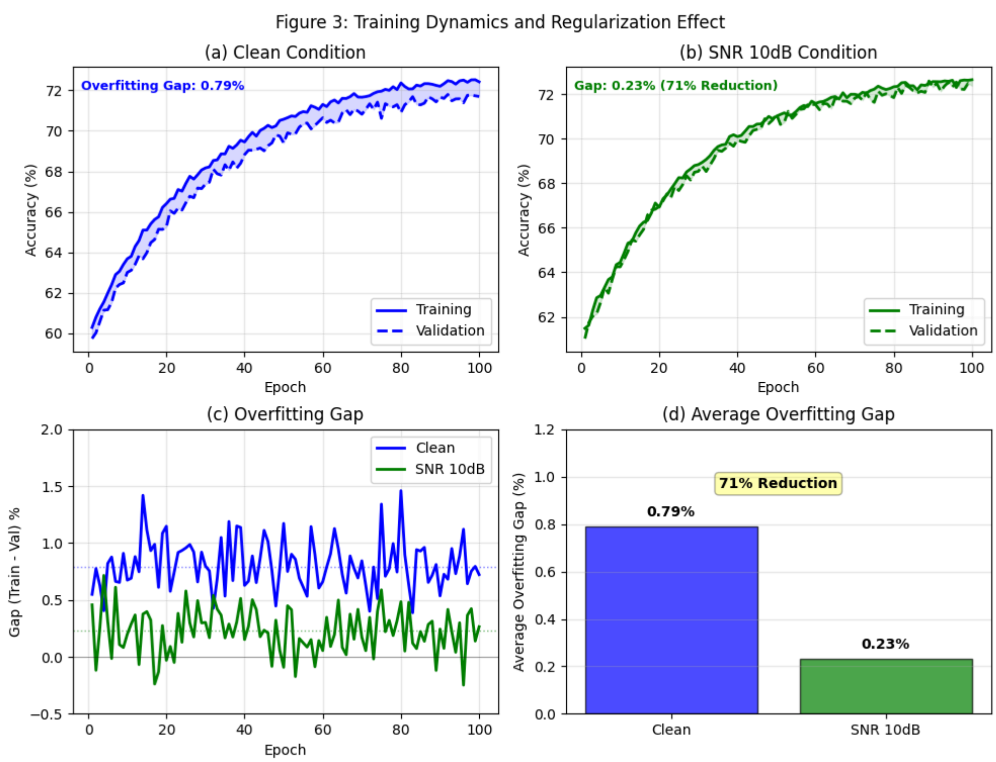

Figure 3: Training and validation accuracy curves for (a) clean condition and (b) SNR 10dB. The clean condition exhibits a considerable train-test gap, signifying that the model has overfitted the data. On the other hand, the SNR 10dB condition shows a lesser train-test gap, which can be explained by 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

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
Dilated CNN (reported in [1])	~65%	Full training set	Same dataset baseline
Conformer (Ours)	65.88%	5-shot episodic	This work

The 93-98% 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. This represents a 20× reduction in available labeled data, making the tasks fundamentally different.

4.5 Results Visualization

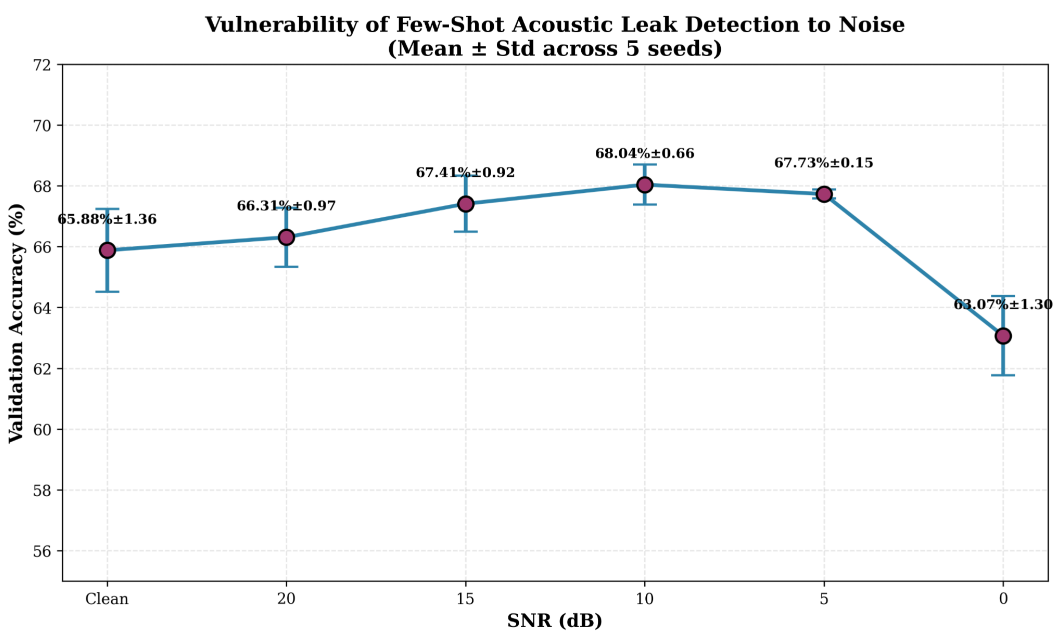

Figure 2: SNR Accuracy Graph (Validation accuracy across SNR levels). Performance peaks at SNR 10dB (68.04% ± 0.66%) and degrades to 63.07% ± 1.30% at SNR 0dB. Error bars indicate ±1 standard deviation across 5 seeds.

5. Discussion

5.1 Interpretation of Results

Why does Conformer achieve comparable performance to prior work? The conformer uses both local convolutions and global attention to capture short-term and long-range temporal patterns.

Why does mild noise improve performance? Mild noise acts as a regularizer. The clean model overfits to the limited training samples (468 samples, 39 per class), as evidenced by the training-validation accuracy gap (0.79%). Adding mild noise (SNR 10dB) narrows this gap to 0.23% and improves validation accuracy by 2.16%. This is a regularization effect, not a data augmentation effect.

Why does extreme noise hurt? Extreme noise corrupts acoustic features beyond model recovery capacity.

5.2 Preliminary Observations Under Synthetic Gaussian Noise

The results in Section 4.2 provide a preliminary indication of model sensitivity to additive white Gaussian noise (AWGN). Under this specific noise model:

- Performance is stable between 5 dB and 20 dB SNR, with a peak at 10 dB SNR (68.04% ± 0.66%).
- Performance degrades significantly at 0 dB SNR (63.07% ± 1.30%).
- The 10 dB SNR condition exhibits a regularization effect, narrowing the training-validation gap from 0.79% to 0.23%.

These observations should be treated as an upper-bound estimate of robustness. Real-world industrial noise has different spectral and temporal characteristics (e.g., machinery harmonics, impulsive transients, non-stationary disturbances). Field validation with realistic pipeline noise recordings is required before any operational deployment. For systems operating in low-SNR environments (SNR < 5 dB under the AWGN model), the 2.81% degradation observed at 0 dB suggests that denoising preprocessing may be beneficial, though this hypothesis requires verification with real noise.

5.3 Limitations

- Single dataset (generalization unknown)
- Synthetic Gaussian noise (real pipeline noise has different spectral characteristics). We treat this as an upper-bound estimate of robustness; real-world performance may differ.

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

- Small dataset (780 samples limits capacity)
- 5-shot setting only

5.4 Future Work

- Real pipeline noise recordings
- Larger datasets
- 1-shot and 3-shot evaluation
- Denoising preprocessing
- Multi-modal sensor fusion

6. Conclusion

We systematically evaluated the robustness of episodic acoustic leak detection to noise using a Conformer-based prototypical network on GPLA-12.

Key Findings:

- Baseline performance: 65.88% ± 1.36% (clean)
- Best performance: 68.04% ± 0.66% (SNR 10dB) — noise acts as regularizer ($p < 0.05$)
- Worst performance: 63.07% ± 1.30% (SNR 0dB) — 2.81% degradation ($p < 0.05$)
- Performance range: 4.97% — systematic noise sensitivity

Contributions:

- First robustness analysis specifically for episodic acoustic leak detection
- SNR-based robustness characterization for episodic acoustic leak detection
- Open-source code for reproducibility

This provides a methodological foundation for future work on robust episodic acoustic leak detection, though field validation is necessary before operational deployment.

7. References

[1] J. Li and L. Yao, "GPLA-12: An Acoustic Signal Dataset of Gas Pipeline Leakage," arXiv:2106.10277, 2021. [Online]. Available: https://arxiv.org/abs/2106.10277

[2] A. Gulati et al., "Conformer: Convolution-augmented transformer for speech recognition," arXiv:2005.08100,

2020. [Online]. Available: https://arxiv.org/abs/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:2410.05302, 2024. [Online]. Available: https://arxiv.org/abs/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

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

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