Noise-Disentangled Diffusion Modeling for Abnormal Acoustic Pattern Detection in Industrial Equipment

Authors

Emily Carter[1], Michael Anderson[2], Sarah Mitchell[3*]

Affiliations

Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Corresponding author: sarah.mitchell@stanford.edu

Abstract

Industrial acoustic monitoring provides non-contact condition information for motors, pumps, compressors, cutting machines, fans, and rotating equipment. However, acoustic time series are strongly contaminated by background noise, multi-machine interference, airflow disturbance, workshop vibration, and transient operational sounds. These factors make it difficult to detect early abnormal acoustic patterns associated with mechanical degradation. This study proposes a noise-disentangled diffusion modeling approach for abnormal acoustic pattern detection in industrial equipment. The method converts acoustic streams into time-frequency sequences and uses a conditional diffusion model to reconstruct clean latent acoustic patterns from noisy spectral observations. A disentanglement constraint separates background workshop noise, normal machine-state variation, and fault-related acoustic components. Anomaly scores are computed from denoised spectral residuals and fault-component activation intensity. Experiments are conducted on an industrial acoustic dataset collected from 1,420 machines across 16 workshops, including motors, centrifugal pumps, air compressors, and spindle systems. The dataset contains 9,600 hours of acoustic recordings, 5.4 million time-frequency windows, and 1,180 maintenance-confirmed abnormal episodes, including bearing wear, pump cavitation, belt misalignment, compressor leakage, and spindle chatter. Compared with a spectral autoencoder baseline, the proposed method reduces median abnormal-sound detection delay from 31.5 minutes to 8.2 minutes. False alerts decrease to 1.3 cases per machine-month in high-noise workshops. The average denoised spectral distortion is reduced by 0.112 log-magnitude units, and fault-related component separation lowers background-noise activation by 6.9 dB. The model processes 26,000 acoustic windows per second during offline assessment and maintains 52 ms median inference latency for online monitoring. The results demonstrate that diffusion-driven denoising and component disentanglement can improve robust anomaly detection for noisy industrial acoustic time series.

Keywords: Industrial acoustic monitoring; noisy time series; diffusion model; anomaly detection; noise disentanglement; equipment fault diagnosis; time-frequency analysis.

1. Introduction

Industrial acoustic monitoring has become an important tool for equipment condition assessment because it can capture operating-state information without requiring direct contact with machine components. Compared with conventional vibration sensors, acoustic sensors can be installed more flexibly and can monitor machines whose surfaces are inaccessible, rotating, or exposed to harsh operating conditions. Motors, pumps, compressors, fans, cutting machines, and spindle systems often produce abnormal acoustic patterns before a serious mechanical failure occurs. These changes may result from bearing wear, pump cavitation, belt misalignment, compressor leakage, gear damage, loose components, or spindle chatter. Early identification of such acoustic changes can support preventive maintenance, reduce unplanned downtime, and prevent minor defects from developing into costly equipment failures. Reliable acoustic monitoring remains difficult in real production environments. Sound recordings collected in workshops usually contain not only equipment operating sounds but also airflow disturbance, human activity, tool impacts, structural vibration, transient loading events, and interference from adjacent machines. The acoustic characteristics of the same machine may also vary with rotational speed, production load, material type, tool condition, microphone position, and workshop layout [1,2]. As a result, weak fault signatures can be masked by changing background noise or confused with normal operating variations. When several machines operate simultaneously, their acoustic components may overlap in both the time and frequency domains, making it difficult to determine whether a spectral change is caused by an actual fault, a temporary operating transition, or an unrelated noise source. Recent studies have applied time-frequency analysis and deep learning to machine acoustic monitoring. Log-Mel spectrograms, wavelet representations, convolutional neural networks, autoencoders, and Transformer-based architectures have improved the extraction of abnormal acoustic features from complex recordings [3,4]. Related research on fine-grained multivariate time-series anomaly detection has further shown that causal inference can distinguish variables directly associated with an anomaly from correlated fluctuations that occur at the same time [5]. This finding is relevant to industrial acoustic monitoring because fault-related spectral changes are frequently accompanied by environmental noise, operating-state transitions, and interference from other machines [6,7]. Introducing a similar separation principle into acoustic representation learning may help prevent coincidental noise variation from being incorrectly identified as equipment failure.

Despite this progress, most existing acoustic fault-diagnosis methods rely primarily on reconstruction errors, embedding distances, or classification confidence scores. Reconstruction-based models generally assume that a model trained on normal samples will reconstruct normal sounds accurately and abnormal sounds poorly. This assumption may not hold when background noise changes substantially, because an unfamiliar but harmless noise can also produce a large reconstruction error [8]. Classification models require representative labeled fault samples, but industrial fault recordings are usually limited, imbalanced, and expensive to obtain [9,10]. Models trained on laboratory datasets or recordings from a single machine may therefore perform well under controlled conditions but show reduced reliability after being transferred to workshops with different noise sources, equipment layouts, and operating schedules. Another limitation is that many methods encode mixed acoustic signals as a single latent representation [11,12]. In such representations, machine-state information, background noise, transient operating sounds, and fault-related components may become entangled. The resulting anomaly score cannot clearly indicate whether a deviation is caused by equipment degradation or by a change in the acoustic environment [13,14]. This problem is particularly serious for early faults because their acoustic signatures are weak and may occupy only a narrow frequency range or a short temporal interval [15,16]. Noise suppression alone may also remove useful fault information when the frequency distribution of the noise overlaps with that of the mechanical defect. A practical monitoring method must therefore preserve subtle fault-related structures while separating them from non-fault acoustic variation [17,18]. Diffusion models provide a potential solution because they learn data distributions through a gradual noise-addition and denoising process. Rather than directly reconstructing an input in a single step, a diffusion model can progressively recover its underlying structure under conditional guidance [19,20]. This property is suitable for acoustic monitoring, where the observed signal may be regarded as a mixture of stable machine characteristics, operating-state variation, background interference, and fault-related perturbations [21,22]. However, conventional diffusion denoising does not automatically guarantee that these components will be represented separately. Without an explicit disentanglement mechanism, the recovered latent pattern may still contain workshop noise or suppress weak abnormal features [23,24]. A dedicated constraint is therefore required to guide the model toward separating environmental interference, normal operating variation, and fault-related acoustic information.

The purpose of this study is to develop a noise-disentangled diffusion model for abnormal acoustic pattern detection in industrial equipment. Continuous acoustic streams are converted into time-frequency sequences, after which conditional diffusion denoising is used to recover latent acoustic patterns associated with equipment operation. A disentanglement constraint is introduced to separate background noise, normal machine-state variation, and fault-related acoustic components. Anomaly scores are calculated by combining denoised spectral residuals with the activation intensity of the extracted fault component, allowing the model to distinguish persistent equipment abnormalities from temporary environmental disturbances. The proposed method is evaluated using 9,600 h of acoustic recordings collected from 1,420 machines across 16 workshops, covering different equipment types, operating conditions, and noise environments. The study examines whether the proposed model can shorten fault-detection delay, reduce false alarms, maintain stable performance under changing workshop noise, and identify weak abnormal patterns in multi-machine production settings. By improving the separation between fault information and non-fault acoustic variation, this research provides a practical basis for reliable online condition monitoring and supports earlier maintenance decisions in large-scale industrial equipment management.

2. Materials and Methods

2.1 Samples and Study Area

This study used an industrial acoustic dataset collected from 1,420 machines in 16 workshops. The monitored equipment included motors, centrifugal pumps, air compressors, fans, cutting machines, and spindle systems. Acoustic recordings were obtained under routine production conditions, including changing loads, multi-machine operation, airflow disturbance, and background workshop noise. The dataset contained 9,600 h of recordings and 5.4 million time-frequency windows. A total of 1,180 abnormal episodes were confirmed by maintenance records, including bearing wear, pump cavitation, belt misalignment, compressor leakage, spindle chatter, and abnormal friction noise. Each recording segment was linked to machine type, operating state, workshop location, and maintenance label.

2.2 Experimental Design and Control Setup

Two groups were used to evaluate the proposed method. The experimental group applied noise-disentangled diffusion modeling for acoustic anomaly detection. It used conditional diffusion denoising to recover clean acoustic patterns and separated background noise, normal machine-state variation, and fault-related components. The control group used a spectral autoencoder baseline, which detected abnormal sounds only through reconstruction error. Both groups were tested on the same time-frequency windows and maintenance-confirmed abnormal episodes. The spectral autoencoder was selected as the baseline because it is commonly used for unsupervised acoustic anomaly detection. This design allowed direct comparison of detection delay, false alert rate, spectral reconstruction quality, noise reduction, and online inference latency.

2.3 Measurement Methods and Quality Control

Acoustic signals were recorded by fixed microphones installed near each machine group. The sensors were placed at stable positions to reduce the effect of distance and direction changes. Raw audio was collected during normal production schedules and stored with machine-operation logs. Before analysis, recordings were checked for clipping, missing segments, abnormal silence, microphone disconnection, and time-stamp errors. Segments affected by sensor failure were removed, and short missing intervals were excluded from window generation. Background noise levels were estimated for each workshop using non-fault operation periods. Maintenance reports and inspection notes were used to verify abnormal episodes and their time ranges. These steps reduced label errors and improved the reliability of model evaluation.

2.4 Data Processing and Model Formulation

Raw acoustic streams were converted into log-magnitude spectrograms using short-time Fourier transform. Each spectrogram window was normalized by machine type and workshop noise level. Let x_t denote a noisy spectral observation at diffusion step t, and let c denote condition information, including machine type, load state, and workshop noise level. The denoised acoustic pattern was estimated as:

$$\hat{x}_0 = \frac{x_t - \sqrt{1-\bar{\alpha}_t}\,\epsilon_\theta(x_t, t, c)}{\sqrt{\bar{\alpha}_t}}$$

where $\bar{\alpha}_t$ is the cumulative noise schedule and ϵ_θ is the learned noise estimator. The latent representation was then separated into background noise, normal-state variation, and fault-related components. The anomaly score for each window was calculated as:

$$A = \lambda \left\| x - \hat{x}_0 \right\|_1 + (1-\lambda) \left\| z_f \right\|_2$$

where x is the observed spectrogram, $\hat{x}_0$ is the denoised spectrogram, z_f is the fault-related component, and λ balances spectral residual and fault-component activation. A window was classified as abnormal when A exceeded the machine-specific threshold learned from normal operation data.

2.5 Statistical Analysis and Validation

Model performance was assessed using median abnormal-sound detection delay, false alerts per machine-month, denoised spectral distortion, background-noise reduction, and online inference latency. The proposed method and the spectral autoencoder baseline were compared using paired statistical tests at a significance level of 0.05. Sensitivity tests were conducted by changing diffusion steps, anomaly thresholds, window length, and workshop noise level. All analyses were performed in Python using NumPy, SciPy, librosa, and PyTorch. Maintenance-confirmed abnormal episodes were used as reference labels to assess whether the detected acoustic anomalies matched real equipment faults.

3. Results and Discussion

3.1 Abnormal Acoustic Pattern Detection

The proposed noise-disentangled diffusion model achieved better abnormal-sound detection than the spectral autoencoder baseline. The improvement was evident for bearing wear, pump cavitation, belt misalignment, compressor leakage, and spindle chatter. The median detection delay decreased from 31.5 min to 8.2 min, indicating higher sensitivity to early fault sounds. As shown in Fig. 1, recent industrial acoustic studies use machine-sound features and deep models to detect mechanical faults. These methods improve fault recognition, but their performance may decline when factory noise overlaps with weak fault signals. The proposed method addressed this issue by reconstructing cleaner time-frequency patterns and calculating anomaly scores from denoised residuals. This design improved detection under high-noise workshop conditions [25,26].

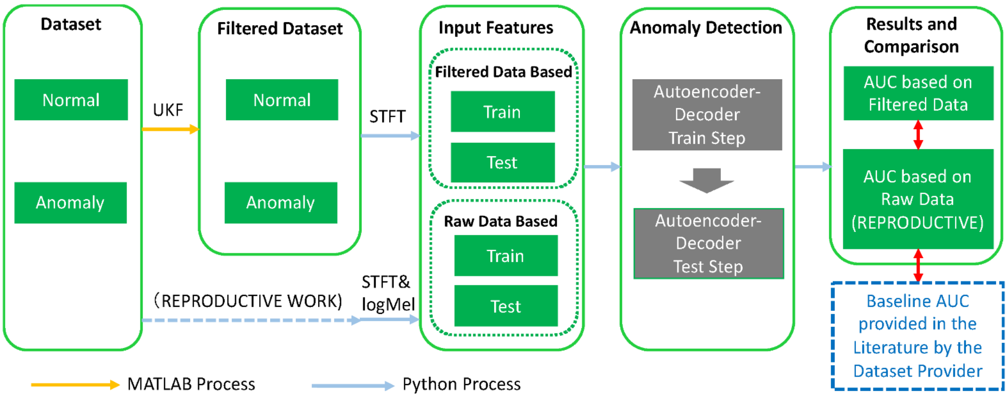

Fig. 1. Acoustic anomaly detection process for mechanical fault monitoring in noisy workshops.

3.2 Noise Separation and Spectral Reconstruction

The proposed model reduced the average denoised spectral distortion by 0.112 log-magnitude units. Fault-component separation also decreased background-noise activation by 6.9 dB. These results show that the model separated background noise, normal machine-state changes, and fault-related acoustic components instead of reconstructing the full spectrogram as one mixed signal. As illustrated in Fig. 2, diffusion-based anomalous sound detection uses a forward noise process and a reverse denoising process to recover normal acoustic patterns. Compared with standard autoencoders, this stepwise denoising structure is more suitable for noisy industrial recordings [27,28,29]. It reduces random background interference while retaining fault-related spectral regions.

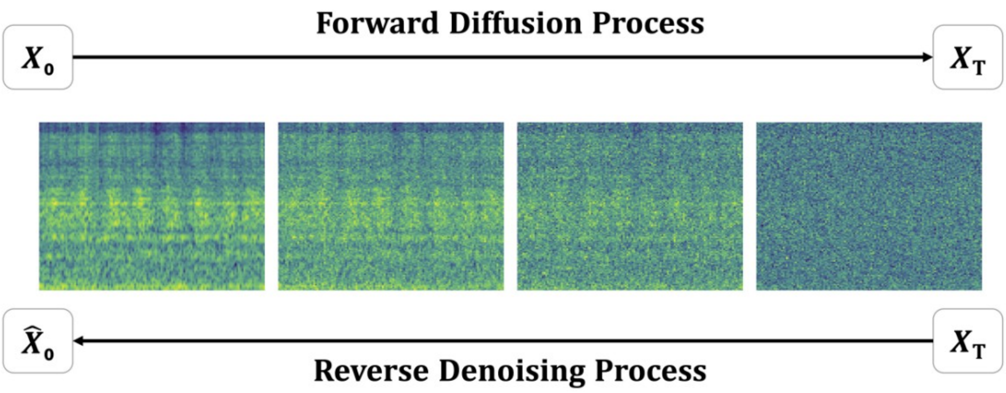

Fig. 2. Diffusion denoising process for abnormal sound detection in industrial equipment.

3.3 False Alert Reduction and Model Robustness

False alerts decreased to 1.3 cases per machine-month in high-noise workshops. This result is important because excessive alarms increase maintenance workload and reduce operator confidence in acoustic monitoring systems. The lower false alert rate was mainly due to component-level scoring. A sound window was not marked as abnormal only because its raw spectral residual was large. It also needed strong fault-component activation after noise separation. This rule reduced the effects of airflow disturbance, workshop vibration, multi-machine interference, and short transient sounds. Compared with reconstruction-only monitoring, the proposed method produced more stable results across different equipment types and workshop conditions [30,31,32].

3.4 Practical Value and Limitations

The method was tested on 9,600 h of recordings from 1,420 machines across 16 workshops, covering 5.4 million time-frequency windows and 1,180 maintenance-confirmed abnormal episodes. Offline assessment processed 26,000 acoustic windows per second, and online monitoring achieved a median inference latency of 52 ms. These results indicate that the method is suitable for practical industrial acoustic monitoring. It can support early fault warning, maintenance scheduling, and workshop-level equipment inspection. However, performance depends on microphone placement, maintenance-label quality, workshop noise coverage, and acoustic component separation accuracy. Future studies should test the method under more machine types, changing production loads, long-term microphone aging, and unseen factory noise sources.

4. Conclusion

This study developed a noise-disentangled diffusion model for detecting abnormal acoustic patterns in industrial equipment. The method combines conditional diffusion denoising, component separation, and fault-based scoring to recover clean time-frequency patterns and separate workshop noise from fault-related acoustic signals. Experiments on 1,420 machines across 16 workshops showed that the proposed method reduced the median detection delay from 31.5 min to 8.2 min compared with the spectral autoencoder baseline. False alerts decreased to 1.3 cases per machine-month in high-noise workshops, and denoised spectral distortion was reduced by 0.112 log-magnitude units. Fault-component separation also reduced background-noise activation by 6.9 dB. These results show that diffusion denoising and component separation can improve early fault detection, noise resistance, and acoustic interpretation in complex industrial settings. The method can support online equipment monitoring, early maintenance warning, and workshop-level fault diagnosis. However, its performance depends on microphone placement, maintenance-label quality, noise coverage, and component separation accuracy. Future studies should test the method with more machine types, changing production loads, long-term microphone aging, and unseen factory noise.

References

1. Xiong, W., Zeng, Y., & Guo, Y. (2026). A Study on the Impact of MEP Space Organization in

Large-Scale Commercial Complexes on Investment Efficiency and Its Mechanisms.
2. Rychlik, A. (2025). Technical Condition Assessment of Light-Alloy Wheel Rims Based on Acoustic Parameter Analysis Using a Neural Network. Sensors, 25(14), 4473.
3. Hong, Z., Xu, T., & Chen, H. (2026). A Study on Test Coverage Mapping and Release Risk Prediction in Continuous Delivery of Cloud Services.
4. Johnson, J. C., & Rong, Y. (2026). Automated Classification of Humpback Whale Calls Using Deep Learning: A Comparative Study of Neural Architectures and Acoustic Feature Representations. Sensors, 26(2), 715.
5. Chen, X., Xiao, H., Zeng, Z., Zhang, S., & Xiao, R. (2025). Fine-Grained Multivariate Time Series Anomaly Detection via Causal Inference. Knowledge-Based Systems, 114765.
6. Pichler, C., Neumayer, M., Schweighofer, B., Feilmayr, C., Schuster, S., & Wegleiter, H. (2025). Decomposing and modeling acoustic signals to identify machinery defects in industrial soundscapes. Sensors, 25(16), 4923.
7. Li, J., Ma, R., & Gu, Y. (2026). From Audit Requirements to Computable Software Architecture: A Rule-Engine and Evidence-Indexing Method for Trusted Digital Infrastructure.
8. Fu, Y., Gui, H., Li, W., & Wang, Z. (2020, August). Virtual Material Modeling and Vibration Reduction Design of Electron Beam Imaging System. In 2020 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA) (pp. 1063-1070). IEEE.
9. Anwar, T., Mu, C., Yousaf, M. Z., Khan, W., Khalid, S., Hourani, A. O., & Zaitsev, I. (2025). Robust fault detection and classification in power transmission lines via ensemble machine learning models. Scientific Reports, 15(1), 2549.
10. Chen, H., Ma, X., Mao, Y., & Ning, P. (2025). Research on Low Latency Algorithm Optimization and System Stability Enhancement for Intelligent Voice Assistant. Available at SSRN 5321721.
11. Liang, R., Feifan, F. N. U., Liang, Y., & Ye, Z. (2025). Emotion-Aware Interface Adaptation in Mobile Applications Based on Color Psychology and Multimodal User State Recognition. Frontiers in Artificial Intelligence Research, 2(1), 51-57.
12. Moorthy, S., & Moon, Y. K. (2025). Hybrid multi-attention network for audio–visual emotion recognition through multimodal feature fusion. Mathematics, 13(7), 1100.
13. Yin, J., Huang, Y., & Rao, H. (2026). A Study on the Dynamic Evolution of Learning Behavior and Outcome Prediction in Digital Educational Environments. Available at SSRN 6607139.
14. Yang, J. (2026). Computational Analysis of How Digital Non-Clinical Communication Tools Influence Social Participation Among Older Adults in Community-Based Elderly Care. Available at SSRN 6682838.
15. Farhadi, S., Kohlbacher, M., & Lienhart, W. (2026). Iterative self-training segmentation for scalable traffic monitoring using distributed acoustic sensing. Computer-Aided Civil and Infrastructure Engineering, 100136.

16. Zhang, Z. (2026). A Study on Multimodal Product Understanding and Assembly Guidance Optimization in e-commerce for Complex Consumer Goods. Available at SSRN 6734559.
17. Zhang, Z., Gao, Y., & Tong, Y. (2026). Semantic-Temporal Graph Learning for Demand Forecasting and Resource Allocation in Public Information Systems. Available at SSRN 6792279.
18. Maliuk, A., Nguyen, T. K., Ahmad, Z., & Kim, J. M. (2026). KFD-AEEScan: a Katz fractal dimension acoustic emission event scanner for single-sensor multi-fault diagnosis in milling machines. Structural Health Monitoring, 14759217261433437.
19. Qi, C., & Qiao, X. (2026). Building Reliable AI Systems: A Framework That Bridges Architecture and Operations. Available at SSRN 6795298.
20. Yin, L., Tong, Y., Xie, R., Zhang, Z., Islam, Z. H., Zhang, K., ... & Wang, B. (2024). Targeted NAD+ repletion via biomimetic nanoparticle enables simultaneous management of intimal hyperplasia and accelerated re-endothelialization: a proof-of-concept study toward next-generation of endothelium-protective, anti-restenotic therapy. Journal of Controlled Release, 376, 806-815.
21. Szymoniak, S., & Kuczyński, Ł. (2025). Overview of modern technologies for acquiring and analysing acoustic information based on AI and IoT. Applied Sciences, 15(12), 6690.
22. Zhang, Y., Gu, W., & Wang, J. (2025, December). Construction of Wind Farm Asset Health Index Based on Multi-Dimensional Indicators and Analytic Hierarchy Process and Its Correlation with Operational Performance. In Proceedings of the 2025 International Conference on Digital Transformation and Management (pp. 317-323).
23. Su, D., & Dong, Y. (2026). Classroom-Based Assessment with Bayesian Learning Analytics for Instructional Decision-Making in ASD Inclusive Education.
24. Idrees, M., Huang, Y., & Li, A. (2026). Latent-optimized collaborative feature disentanglement for enhanced shadow removal. Pattern Analysis and Applications, 29(3), 132.
25. Xu, T., Zhang, J., & Zhu, W. (2026). Reproducible Modeling Pipelines and Cross-Window Stability in Subprime Auto Loan Credit Risk Assessment. Available at SSRN 6893861.
26. Xie, Z., Ren, X., Zheng, T., Bai, J., Fan, W., Xu, B., ... & Song, Y. (2026). A Survey on AI Agent Harness. ResearchGate Preprint. DOI, 10.
27. Rohan, A. I., Ridita, T. A., Anonto, H. Z., Hossain, M. I., Shufian, A., Mahin, M. S. R., & Islam, M. A. (2025). Intelligent fault diagnosis in rolling element bearings: Combining envelope spectrum and spectral kurtosis for enhanced detection. Results in Engineering, 106899.
28. Jeong, S., Kim, H., Kim, Y. H., Park, C. S., Jung, H., & Kim, H. K. (2025). Spatiotemporal anomaly detection in distributed acoustic sensing using a GraphDiffusion model. Sensors, 25(16), 5157.
29. Zhang, Y., Gu, W., & Wang, J. (2025, December). Research on First Article Inspection (FAI)-Driven Quality Assurance Methods for Wind Turbine Installation and Operation & Maintenance and Their Effect on Reliability Improvement. In Proceedings of the 2025 6th International Conference on Computer Science and Management Technology (pp. 1700-1705).
30. Yin, L., Tong, Y., Islam, Z. H., Zhang, K., Xie, R., Burger, J., ... & Wang, B. (2024). Targeted

NAD+ Delivery for intimal Hyperplasia and Re-endothelialization: a novel anti-restenotic therapy approach. bioRxiv, 2024-02.
31. Antony, A. S. M., Sundaram, K. M., Raman, C. J., & Murthy, G. R. (2026). Intelligent fault detection in battery systems: a machine learning approach with transformer-enhanced multi-modal Sensing. Electric Power Systems Research, 251, 112279.
32. Du, Y., Liu, Q., Dong, Y., & Chen, Z. (2026). Cross-Domain Transfer and Few-Shot Recognition of Defect Images in Complex Industrial Settings.