ABNORMAL SOUND DETECTION BASED ON NOISE-AWARE DOMAIN GENERALIZATION

Technical Report

Zhifang Zheng, Shengbing Chen, Yanjun Zhou, Younan Ji, Hanbin Zhou, Shuchi Chen

School of Artificial Intelligence and Big Data
Hefei University, Hefei, China
281351712@qq.com,shbchen@hfuu.edu.cn,1554024134@qq.com,2740819308@qq.com,
2384883301@qq.com, 2929950149@qq.com

ABSTRACT

This technical report describes our anomalous sound detection systems submitted for DCASE 2026 Challenge Task 2. The task focuses on noise-aware unsupervised anomalous sound detection for machine condition monitoring, where only normal machine sounds are available for training and the test data may include domain shifts and environmental noise. To address this setting, we build several simple systems based on pre-trained audio representations, two-channel noise-aware auxiliary features, and distance-based anomaly scoring. The submitted systems mainly differ in the combination of BEATs embeddings, EAT embeddings, and spatial features extracted from near- and far-microphone recordings. For each machine type, normal training samples are used as reference data, and anomaly scores are obtained by measuring the distance between test samples and the reference normal samples. The final submitted systems are selected on the development set according to the official metrics.

Index Terms— DCASE, anomalous sound detection, machine condition monitoring, domain generalization, noise-aware learning

1. INTRODUCTION

This report describes the abnormal sound detection method developed for DCASE 2026 Challenge Task 2[1] Anomalous sound detection (ASD) aims to identify whether the sound emitted from a target machine is normal or anomalous. Since anomalous machine sounds are difficult and costly to collect in real industrial environments, the task is formulated as an unsupervised anomaly detection problem, in which only normal sounds are provided for training.

Recent DCASE Task 2 settings have emphasized first-shot deployment and domain generalization[2],[3] In such scenarios, the acoustic characteristics of the development and evaluation machines may be different, and the system should work without machine-specific manual tuning. DCASE 2026 further introduces a noise-aware setting. Each sample contains two-channel recordings captured near and far from the target machine. The far-channel signal is expected to contain stronger environmental noise, while the near-channel signal contains relatively clearer machine sounds. This two-channel design provides useful information for robust machine condition monitoring under noisy conditions[1].

Our submitted systems follow a practical and compact design. We use pre-trained audio models to extract transferable sound representations, and use the near- and far-channel relationship to obtain auxiliary noise-aware features. A distance-based anomaly detection backend is then applied to normal reference samples. Four system variants are submitted by changing the combination of feature branches and score-level fusion strategies.

2. EXPERIMENTAL SETUP

2.1. Dataset

The systems are developed on the DCASE 2026 Task 2 development dataset. The development dataset contains seven machine types: ToyCarEmu, ToyCar, bearingEmu, fan, gearboxEmu, sliderEmu, and valveEmu. Each machine type contains normal training samples and test samples from source and target domains. The evaluation dataset contains five different machine types: BlowerDustCollector, Sander, SewingMachine, ToothBrush, and ToyDrone. The labels of the evaluation dataset are not used during system development.

2.2. Features Processing

All audio signals are resampled to 16 kHz. The two channels are both retained. The development dataset is used for checking the system configuration and selecting the submitted variants. No external anomalous sound dataset is used. The external resources used in this work are pre-trained audio models, including BEATs[5] and EAT[6].

The official metrics include the area under the receiver operating characteristic curve (AUC) and partial AUC (pAUC). AUC is calculated for source and target domains, and pAUC is calculated under the official false-positive-rate range. The final score is based on the harmonic mean of the official metric terms over all machine types.

Table 1: Anomaly detection results for different machine types

Model	Score	ToyCar(Emu)	ToyCar	Bearing	Fan	GearBox(Emu)	Slider(Emu)	Valve(EMu)
Baseline(Mahala)	*s_AUC*	69.49	77.28	65.92	60.00	74.48	66.36	56.60
	t_AUC	66.62	53.17	62.28	45.09	52.74	49.18	56.50
	pAUC	53.47	58.25	60.42	52.29	53.97	50.36	50.20
Method1	*s_AUC*	70.36	83.10	61.06	83.90	78.58	60.66	82.48
	t_AUC	63.38	66.66	60.10	51.06	66.40	49.66	87.08
	pAUC	55.37	62.26	59.53	50.47	55.05	48.95	62.32
Method2	*s_AUC*	71.40	85.02	58.52	78.72	79.16	63.34	82.56
	t_AUC	62.94	64.12	57.60	49.58	66.22	49.86	88.58
	pAUC	55.26	63.37	58.79	51.74	58.05	49.05	65.00
Method3	*s_AUC*	70.24	82.48	61.08	83.52	78.42	60.36	82.42
	t_AUC	61.28	67.28	60.08	51.50	66.20	49.92	87.10
	pAUC	55.00	62.47	60.11	50.84	54.79	48.84	61.95
Method4	*s_AUC*	71.12	84.36	58.58	77.98	79.26	63.20	82.56
	t_AUC	61.58	64.64	57.56	50.26	65.64	50.16	88.64
	pAUC	54.89	62.21	59.00	51.95	57.79	48.84	64.74

3. PROPOSED METHOD

Pre-trained audio representation. Large-scale self-supervised audio models provide general acoustic representations that are useful for downstream sound analysis. In our system, BEATs is used as the main feature extractor[5]. BEATs takes waveform input and outputs frame-level audio embeddings. A lightweight adaptation strategy is applied to the feature extractor so that the representation is better matched to machine sounds. This adaptation is implemented in a parameter-efficient manner, following the general idea of low-rank adaptation[7]. Some submitted systems additionally use EAT embeddings[6] to increase representation diversity.

For each audio clip, frame-level embeddings are converted to clip-level representations by simple temporal statistics. We mainly use mean and standard deviation pooling. This setting keeps the backend simple and avoids introducing a complex trainable classifier on top of the pre-trained representations.

Noise-aware spatial feature. DCASE 2026 Task 2 provides near- and far-microphone signals. We use this two-channel structure to extract auxiliary noise-aware information. The general idea is to compare the two channels and construct statistical features that reflect the difference between machine-dominant and noise-dominant recordings. This feature branch is used as a complement to the pre-trained audio embeddings.

Anomaly detection backend. The backend follows a reference-based anomaly detection scheme. For each machine type, normal training samples are stored as a reference bank in the feature space. During testing, the anomaly score is obtained by measuring the distance between the test sample and the reference normal samples. This is related to classical distance-based and nearest-neighbor outlier detection methods[8]. A larger distance indicates that the test sample is less consistent with the normal training distribution.

The anomaly scores from different feature branches are normalized for each machine type and then combined at the score level. We do not use anomalous samples for model training. The final submitted systems are selected according to the development-set performance.

4. RESULTS

The detailed results of the submitted systems are summarized in the result table. We report the official AUC, pAUC, and final harmonic-mean score on the development set. Machine-wise results are also included to show the performance difference across machine types. In general, the systems combining pre-trained audio embeddings with two-channel auxiliary features are more stable than the baseline-style configuration.

The submitted systems show different behavior across machine types. This suggests that both domain shift and environmental noise remain important factors in the task. The final submitted variants are selected by considering the overall official score and the stability of the system across source and target domains.

5. CONCLUSION

This technical report presents our systems for DCASE 2026 Challenge Task 2. The systems are based on pre-trained audio embeddings, two-channel noise-aware auxiliary features, and distance-based anomaly detection. Four variants are submitted by changing the feature combination and score-level fusion strategy. The system design is simple and does not use anomalous samples for training, making it suitable for the unsupervised and noise-aware setting of the challenge.

6. REFERENCES

[1] T. Nishida, N. Harada, D. Takeuchi, D. Niizumi, K. Imoto, K. Dohi, H. Purohit, T. Endo, and Y. Kawaguchi, "Description and discussion on DCASE 2026 Challenge Task 2: Noise-aware unsupervised anomalous sound detection for machine condition monitoring," arXiv:2606.01578, 2026.

[2] T. Nishida, N. Harada, D. Niizumi, D. Albertini, R. Sannino, S. Pradolini, F. Augusti, K. Imoto, K. Dohi, H. Purohit, T. Endo, and Y. Kawaguchi, "Description and discussion on DCASE 2025 Challenge Task 2: First-shot unsupervised anomalous sound detection for machine condition monitoring," arXiv:2506.10097, 2025.

[3] T. Nishida, N. Harada, D. Niizumi, D. Albertini, R. Sannino, S. Pradolini, F. Augusti, K. Imoto, K. Dohi, H. Purohit, T. Endo, and Y. Kawaguchi, "Description and discussion on DCASE 2024 Challenge Task 2: First-shot unsupervised anomalous sound detection for machine condition monitoring," arXiv:2406.07250, 2024.

[4] Y. Kawaguchi, K. Imoto, Y. Koizumi, N. Harada, D. Niizumi, K. Dohi, R. Tanabe, H. Purohit, and T. Endo, "Description and discussion on DCASE 2021 Challenge Task 2: Unsupervised anomalous detection for machine condition monitoring under domain shifted conditions," in Proc. DCASE Workshop, pp. 186-190, 2021.

[5] S. Chen, Y. Wu, C. Wang, S. Liu, D. Tompkins, Z. Chen, W. Che, X. Yu, and F. Wei, "BEATs: Audio pre-training with acoustic tokenizers," in Proc. ICML, 2023.

[6] W. Chen, Y. Liang, Z. Ma, Z. Zheng, and X. Chen, "EAT: Self-supervised pre-training with efficient audio transformer," in Proc. IJCAI, pp. 3807-3815, 2024.

[7] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, "LoRA: Low-rank adaptation of large language models," in Proc. ICLR, 2022.

[8] S. Ramaswamy, R. Rastogi, and K. Shim, "Efficient algorithms for mining outliers from large data sets," in Proc. ACM SIGMOD, pp. 427-438, 2000.