EAT-LORA AND NEAR/FAR STATISTICAL FEATURES FOR DCASE 2026 TASK 2 ANOMALOUS SOUND DETECTION

Technical Report

Baohe Zhu, LiLi Zhang, Ding Li
ehoab@qq.com, lilizhang@gmail.com, dingli@gmail.com
Shanghai Funda Acoustics Engineering Co., Ltd., China

ABSTRACT

This report describes the FDA submission to DCASE 2026 Task 2. The submitted systems are fixed-artifact, normal-only anomalous sound detectors designed for machine condition monitoring with two-channel recordings. Each system combines a handcrafted statistical branch using near/far acoustic information with an Efficient Audio Transformer (EAT) embedding branch adapted by parameter-efficient LoRA training. For each branch, train-normal examples are used to fit robust normalization, PCA, Mahalanobis, and k-nearest-neighbor scoring artifacts. A domain-aware alpha adjustment, optional score clipping, and equal-weight branch fusion produce the final anomaly score. The binary decision threshold is the 0.99 quantile of train-normal fused scores and is not used for ranking metrics. On the development set, the best submitted configuration obtains 62.83 percent Omega.

***Index Terms*—** EAT, LoRA, Mahalanobis distance, k-nearest neighbors, score fusion.

1. INTRODUCTION

The dataset of DCASE 2026 Task 2 inherits previous Task 2 designs built upon ToyADMOS2 [1] and MIMII DG [2] for normal-only machine anomalous sound detection under domain and noise mismatch [3], [4]. The official autoencoder/Selective Mahalanobis baseline follows a first-shot domain-generalization system for machine condition monitoring [5]. The available two-channel near/far recordings provide complementary information: a near microphone captures the target machine more strongly, while a far microphone can expose background and environmental components. Our design combines a transparent two-channel statistical detector with a representation-learning branch based on EAT embeddings.

The guiding principle is to avoid evaluation-set adaptation. All parameters are fixed before evaluation inference. The development split is used only to report metrics and compare fixed systems, while the additional-training split is used for fitting final artifacts. The submitted anomaly score is continuous and larger-is-more-anomalous. A decision threshold is provided for the official binary decision file, but ranking metrics are determined only by anomaly scores.

2. PROPOSED METHOD

Audio front end and statistical branch. Each recording is treated as a 16 kHz stereo signal with near and far channels. The statistical branch extracts compact descriptors from the two channels, including log-mel summary statistics and near/far difference features. The submitted systems use either absolute near/far-difference statistics or current log-mel summary statistics as the statistical feature group. These interpretable features complement transformer embeddings and remain stable when training data are limited.

EAT-LoRA embedding branch. The second branch uses an allowed pre-trained Efficient Audio Transformer checkpoint as an audio representation model [6]. Variable-length masked inference prevents padded frames from contributing to pooled representations. Long recordings are chunked when necessary, and chunk embeddings are aggregated by valid-token masked pooling. LoRA modules adapt attention-related query/value projections while the backbone remains frozen [7]. The classifier is used only as a proxy normal-data training objective; the normalized projection embedding is used for anomaly scoring.

Normal-only scoring. For each branch and machine/section group, only train-normal examples fit scoring artifacts. The pipeline applies robust train-normal score normalization, PCA with 0.99 retained variance, diagonal Mahalanobis distance [8], and k-nearest-neighbor distance with k=1 and cosine distance [9]. During train-score normalization the self-neighbor is excluded, and no evaluation score distribution is used to refit normalizers or thresholds.

Domain-aware adjustment and score fusion. The raw branch score is adjusted using a source/target prototype distance ratio. Systems 1 and 3 use alpha=0.8, system 2 uses alpha=1.0, and system 4 disables the adjustment. Systems 1 to 3 clip adjusted branch scores at 10. The final anomaly score is $S(x)=0.5\ S_{stat}(x)+0.5\ S_{EAT}(x)$. The submitted decision result uses a fixed train-normal quantile threshold q=0.99; AUC, pAUC, and Omega are ranking metrics based on S(x) [3], [4].

3. SUBMITTED SYSTEMS

S1-S4 correspond to submission labels Zhu_FDA_task2_1 through Zhu_FDA_task2_4, respectively. The fixed systems differ only in statistical feature group, LoRA size, alpha, and clipping; no dense search, evaluation labels, evaluation score-distribution tuning, or per-machine/per-domain/per-section weights are used.

ID	Submission	Stat. group	EAT LoRA	Alpha	Clip	Fusion	Dev. Omega (%)
S1	Zhu_FDA_task2_1	Near/far difference	r8/a16	0.8	10	0.5/0.5	62.83
S2	Zhu_FDA_task2_2	Near/far difference	r8/a16	1.0	10	0.5/0.5	61.91
S3	Zhu_FDA_task2_3	Current mel statistics	r8/a16	0.8	10	0.5/0.5	61.01
S4	Zhu_FDA_task2_4	Near/far difference	r4/a8	none	none	0.5/0.5	61.60

Table 1: Configuration of the four submitted systems.

ID	AUC_source HM	AUC_target HM	pAUC HM	Omega	Observation
S1	73.01	60.11	57.42	62.83	Best balance
S2	73.73	57.10	57.54	61.91	Higher source AUC
S3	72.82	57.02	55.87	61.01	Mel-stat variant
S4	73.48	57.36	56.61	61.60	Small LoRA reference

Table 2: Development-set results in percent. Overall values are harmonic means across development machine types.

4. EXPERIMENTAL SETUP

Development experiments use ToyCarEmu, ToyCar, bearingEmu, fan, gearboxEmu, sliderEmu, and valveEmu. The train split fits normal-only artifacts, while the development test split is used only for source-domain AUC, target-domain AUC, pAUC, and Omega. Overall values are percent harmonic means, penalizing low performance on any domain, machine, or low-false-positive region.

5. RESULTS AND DISCUSSION

Zhu_FDA_task2_1 is selected as the primary configuration because it gives the highest development Omega and the best balance between source-domain AUC, target-domain AUC, and pAUC. Zhu_FDA_task2_2 increases the source-domain harmonic mean but decreases the target-domain harmonic mean, suggesting that alpha=1.0 is less conservative for domain-shifted normal data. Zhu_FDA_task2_3 remains competitive but has lower pAUC and Omega, indicating that absolute near/far difference features are useful for low-false-positive operation. Zhu_FDA_task2_4 uses a smaller LoRA adapter and no alpha/clip adjustment; its result is close to systems 2 and 3, but the best development balance is obtained by the r8/a16 LoRA system with alpha=0.8 and clipping.

6. CONCLUSION

We submitted four fixed systems for DCASE 2026 Task 2. The method combines two-channel statistical near/far features with EAT-LoRA embeddings and normal-only Mahalanobis/kNN scoring. The best development result is obtained by the r8/a16 EAT-LoRA system using absolute near/far-difference statistical features, alpha=0.8, clipping, and equal branch fusion. The design emphasizes reproducibility and avoids evaluation-set tuning. Future work will focus on explicit noise modeling and more adaptive source-target calibration while keeping the evaluation protocol fixed.

7. REFERENCES

[1] N. Harada, D. Niizumi, D. Takeuchi, Y. Ohishi, M. Yasuda, and S. Saito, "ToyADMOS2: Another dataset of miniature-machine operating sounds for anomalous sound detection under domain shift conditions," in Proc. DCASE Workshop, Barcelona, Spain, Nov. 2021, pp. 1-5.

[2] K. Dohi, T. Nishida, H. Purohit, R. Tanabe, T. Endo, M. Yamamoto, Y. Nikaido, and Y. Kawaguchi, "MIMII DG: Sound dataset for malfunctioning industrial machine investigation and inspection for domain generalization task," in Proc. 7th DCASE Workshop (DCASE2022), Nancy, France, Nov. 2022.

[3] DCASE Community, "DCASE 2026 Challenge Task 2: Noise-aware Unsupervised Anomalous Sound Detection for Machine Condition Monitoring,".

[4] Tomoya Nishida, Noboru Harada, Daiki Takeuchi, Daisuke Niizumi, Keisuke Imoto, Kota Dohi, Harsh Purohit, Takashi Endo, and Yohei Kawaguchi. Description and discussion on DCASE 2026 challenge task 2: noise-aware unsupervised anomalous sound detection for machine condition monitoring. In arXiv e-prints: 2606.01578, 2026.

[5] N. Harada, D. Niizumi, D. Takeuchi, Y. Ohishi, and M. Yasuda, "First-shot anomaly detection for machine condition monitoring: A domain generalization baseline," in Proc. 31st European Signal Processing Conference (EUSIPCO), 2023, pp. 191-195.

[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] P. C. Mahalanobis, "On the generalized distance in statistics," Proceedings of the National Institute of Sciences of India, vol. 2, pp. 49-55, 1936.

[9] T. Cover and P. Hart, "Nearest neighbor pattern classification," IEEE Transactions on Information Theory, vol. 13, no. 1, pp. 21-27, 1967.