THUEE SYSTEM FOR DCASE 2026 TASK 2: ENSEMBLING MULTIPLE ANOMALY DETECTORS FOR GENERALIZED ANOMALOUS SOUND DETECTION

Technical Report

Shuwei Zhang[1], Wenrui Liang[1], Xinhu Zheng[2], Lvxin Xu[1], Anbai Jiang[1], Tianyu Liu[1]*
Pingyi Fan[1], Wei-Qiang Zhang[1], Yanmin Qian[2], Xie Chen[2], Cheng Lu[3], Jia Liu[1,4]

[1] Tsinghua University, Beijing, China
[2] Shanghai Jiao Tong University, Shanghai, China
[3] North China Electric Power University, Beijing, China
[4] Huakong AI Plus Company Limited, Beijing, China
*Email: zhangsw25@mails.tsinghua.edu.cn

ABSTRACT

This report presents the THUEE submission to the DCASE 2026 Challenge Task 2, which addresses noise-aware anomalous sound detection (ASD) for machine condition monitoring. The task introduces a dual-channel setup where one microphone captures near-field machine sound and the other captures far-field, noise-contaminated recordings. Our system builds on an audio foundation model, OpenBEATs, and fine-tunes it with a joint classification objective over machine type, operating attribute, and channel index. To enhance generalization, we augment the training data with pseudo spectrograms generated by a generative model. For anomaly detection, we extract utterance-level embeddings and employ five diverse scoring strategies: three KNN-based detectors (including local density rescaling and VarMin), a Mahalanobis-distance-based detector, and a flow-matching-based detector. Anomaly scores from these detectors are fused through weighted linear combination optimized via Bayesian search. Four system variants are submitted, among which the ensemble combining KNN, Mahalanobis, and flow matching detectors achieves the best overall harmonic mean of 69.54% on the development set. We further observe that although the far-field channel benefits representation learning during fine-tuning, it tends to introduce distraction during anomaly scoring, highlighting the need for careful channel-level treatment.

Index Terms— Anomalous Sound Detection, Audio Foundation Model, KNN, Flow Matching

1. INTRODUCTION

Anomalous Sound Detection (ASD) has recently been dominated by audio foundation models, as evidenced by a series of influential works [1, 2, 3, 4]. These studies typically fine-tune powerful pretrained backbones—such as BEATs [5] and EAT [6]—on machine audio via classification over working-condition attributes, and subsequently extract embeddings for KNN-based anomaly detection. This paradigm has been widely validated in numerous DCASE challenge submissions [7, 8] and peer-reviewed publications [9, 10]. Following this established direction, the THUEE system for the DCASE 2026 challenge is also built upon audio foundation models.

The DCASE 2026 Challenge Task 2 [11] centers on noise-aware anomalous sound detection for machine condition monitoring. While the task builds upon earlier machine-sound datasets and domain-generalization benchmarks—including ToyADMOS2 [12], MIMII DG [13], and first-shot anomaly detection for machine condition monitoring [14]—it introduces a key distinction: a deliberate emphasis on dual-channel modeling under realistic noise conditions. All recordings are provided in two channels, where channel 1 captures sound in close proximity to the target machine, while channel 2 is recorded at a greater distance and thus contains higher levels of ambient noise. This dual-channel setup offers complementary information about the machine's operating status, yet also necessitates careful design to fully exploit the semantic richness embedded across both channels.

Building on our prior work [8], this submission remains firmly grounded in audio foundation models. Specifically, we adopt OpenBEATs [15] as our backbone and fine-tune it on the DCASE 2026 dataset through classification over the joint space of machine type, operating attribute, and channel index. To further enrich the training data, we synthesize pseudo spectrograms using a generative model for data augmentation [16]. For the detection phase, we explore five distinct anomaly scoring strategies, including three KNN-based variants, one Mahalanobis-distance-based method, and one flow-matching-based detector. These heterogeneous designs are ultimately integrated through a score fusion mechanism, yielding four comprehensive systems. Among them, our best-performing system achieves an overall score of 69.54% on the development set. We also observe an intriguing finding: while channel 2 proves beneficial for model fine-tuning, it tends to introduce distraction during anomaly detection, highlighting the need for careful channel-level treatment.

2. SYSTEM DESCRIPTION

The THUEE system is built entirely on an audio foundation model, i.e. OpenBEATs[15]. We first fine-tune the model on the DCASE 2026 dataset, then set up multiple anomaly detectors, and finally fuse their scores into powerful ensembles. The system design is described as follows.

Table 1: Development-set Results (%) of the THUEE Submission

Machine	Metric	System 1	System 2	System 3	System 4
bearingEmu	AUC_s	63.90	67.16	67.44	65.40
	AUC_t	62.24	64.12	65.56	66.38
	pAUC	60.11	61.37	61.16	61.21
	hmean	62.04	64.13	64.61	64.25
fan	AUC_s	82.94	89.30	88.78	82.86
	AUC_t	50.94	71.58	69.68	56.62
	pAUC	55.21	64.79	63.26	58.11
	hmean	60.24	73.89	72.43	63.91
gearboxEmu	AUC_s	85.04	80.04	80.34	81.90
	AUC_t	73.22	83.08	82.40	76.84
	pAUC	66.68	69.42	67.16	62.79
	hmean	74.23	77.05	76.00	72.90
sliderEmu	AUC_s	68.04	62.60	62.40	64.96
	AUC_t	61.98	66.34	67.46	65.78
	pAUC	51.58	51.89	51.95	52.00
	hmean	59.74	59.62	59.88	60.21
ToyCar	AUC_s	76.34	71.32	74.56	79.00
	AUC_t	75.72	79.86	82.06	83.26
	pAUC	64.63	63.63	65.68	70.84
	hmean	71.81	70.99	73.49	77.35
ToyCarEmu	AUC_s	72.08	72.68	71.14	62.30
	AUC_t	94.14	94.00	92.82	90.96
	pAUC	66.53	54.53	54.00	53.58
	hmean	75.90	70.20	69.21	65.63
valveEmu	AUC_s	70.22	73.12	75.36	76.76
	AUC_t	73.38	84.84	85.82	81.92
	pAUC	55.37	64.79	64.58	58.05
	hmean	65.32	73.35	74.24	70.65
Overall	AUC_s	74.65	75.17	75.72	73.31
	AUC_t	70.80	77.69	77.40	74.82
	pAUC	60.02	61.49	61.11	59.28
	hmean	66.45	69.41	**69.54**	67.40

Overall hmean values follow the submitted system manifest. Overall AUC_s, AUC_t, and pAUC are arithmetic averages over machine types.

2.1. Model Fine-tuning

We fine-tune OpenBEATs on the DCASE 2026 dataset to learn discriminative and high-fidelity representations tailored for downstream anomaly detection. OpenBEATs is a reproduced and enhanced variant of BEATs [5], which surpasses the original by leveraging larger-scale pre-training data. Our preliminary experiments indicate that OpenBEATs exhibits superior adaptability to the ASD downstream task. We adopt the iter3 checkpoint of OpenBEATs as our backbone. Following the design of AnoPatch [1], we equip the backbone with an attentive statistical pooling layer [17] and a linear classification head. Given that OpenBEATs is pre-trained on mono-channel audio, we split each recording from the DCASE 2026 dataset into two mono-channel clips. We then treat every unique combination of machine type, operating condition attribute, and channel index as a distinct class for classification, and fine-tune the model under these class labels. Our early results show that treating the two channels as separate classes provides stronger supervisory signals than using a single channel alone. The model is fine-tuned using parameter-efficient fine-tuning (PEFT) strategies, following the approach in [9]. Additionally, we incorporate a generative model to synthesize pseudo spectrograms, thereby augmenting the training set and enhancing the model's generalization capability [16].

2.2. Anomaly Detection

After the fine-tuning process, the model stands as an informative and powerful backbone for feature extraction. Thus, we extract its utterance-level embeddings and conduct anomaly detection purely on these embeddings. To fully exploit the latent semantics within these embeddings, we employ five distinct anomaly detectors:

- **Vanilla KNN detector**: This detector is identical to the detector of AnoPatch, which consists of two detectors, each for a domain. Both detectors calculate the cosine distance to the nearest neighbor (k=1) in the memory banks, and the minimum of the two distances is utilized as the final anomaly score.

- **KNN detector with local density rescaling**: We improve the vanilla KNN detector with local density rescaling [18].
- **KNN detector with local density rescaling and VarMin**: We further improve the rescaled KNN detector with VarMin [19].
- **Mahalanobis detector**: This detector assumes that normal embeddings follow a Gaussian distribution for each machine type. It estimates the mean and covariance of the training embeddings and calculates the Mahalanobis distance of a query embedding.
- **Flow Matching detector**: We explore a novel anomaly detector based on flow matching. The flow matching model is trained on normal embeddings and leverages the flow mismatch to detect anomalies [20]. This detector is similar to reconstruction-based methods [21]. However, it is much more accurate and robust.

2.3. Score Fusion

We fuse the anomaly scores of different anomaly detectors by linear combination, where the coefficients are searched by Bayes optimization on the development set of the DCASE 2026 dataset.

3. SUBMITTED SYSTEMS

We submit four systems:

- **System 1**: a single flow matching scoring system.
- **System 2**: an ensemble system that combines KNN and Mahalanobis detectors.
- **System 3**: an ensemble system that combines KNN, Mahalanobis, and flow matching detectors.
- **System 4**: an ensemble system that combines KNN and flow matching detectors.

4. EXPERIMENT RESULTS

Detection performance is evaluated using source-domain AUC (AUC_s), target-domain AUC (AUC_t), partial AUC (pAUC), and their harmonic mean, following the challenge rules. Table 1 reports the development-set performance of the four submitted systems.

5. CONCLUSION

In this work, we presented the THUEE system for the DCASE 2026 Challenge Task 2 on noise-aware anomalous sound detection. Built upon the OpenBEATs audio foundation model, our approach leverages joint classification of machine type, operating attribute, and channel index to learn discriminative representations, and employs generative pseudo-spectrogram augmentation to improve generalization. We investigated five complementary anomaly scoring strategies, including three KNN-based detectors, a Mahalanobis-distance-based detector, and a flow-matching-based detector, and combined them through linearly weighted score fusion optimized on the development set. Among the four submitted systems, the ensemble incorporating KNN, Mahalanobis, and flow matching detectors achieved the best overall harmonic mean of 69.54%. Our analysis further reveals a nuanced channel-level effect: while the far-field channel provides valuable supervisory signals during fine-tuning, it tends to introduce distracting information during anomaly detection, motivating more sophisticated channel-aware designs.

6. REFERENCES

[1] A. Jiang, B. Han, Z. Lv, Y. Deng, W.-Q. Zhang, X. Chen, Y. Qian, J. Liu, and P. Fan, "Anopatch: Towards better consistency in machine anomalous sound detection," in *Interspeech 2024*, 2024, pp. 107–111.

[2] Z. Lv, A. Jiang, B. Han, Y. Liang, Y. Qian, X. Chen, J. Liu, and P. Fan, "Aithu system for first-shot unsupervised anomalous sound detection," DCASE2024 Challenge, Tech. Rep., June 2024.

[3] X. Zheng, A. Jiang, B. Han, Y. Qian, P. Fan, J. Liu, and W.-Q. Zhang, "Improving anomalous sound detection via low-rank adaptation fine-tuning of pre-trained audio models," in *2024 IEEE Spoken Language Technology Workshop (SLT)*. IEEE, 2024, pp. 969–974.

[4] P. Fan, A. Jiang, S. Zhang, X. Zheng, Z. Lv, B. Han, W. Liang, J. Li, W.-Q. Zhang, Y. Qian, X. Chen, and J. Liu, "Fisher: A foundation model for multimodal industrial signal comprehensive representation," *IEEE Transactions on Industrial Informatics*, pp. 1–12, 2026.

[5] S. Chen, Y. Wu, C. Wang, S. Liu, D. Tompkins, Z. Chen, and F. Wei, "Beats: Audio pre-training with acoustic tokenizers," *arXiv preprint arXiv:2212.09058*, 2022.

[6] W. Chen, Y. Liang, Z. Ma, Z. Zheng, and X. Chen, "Eat: Self-supervised pre-training with efficient audio transformer," *arXiv preprint arXiv:2401.03497*, 2024.

[7] X. Zheng, A. Jiang, B. Han, S. Zhang, W.-Q. Zhang, X. Chen, C. Lu, P. Fan, J. Liu, and Y. Qian, "Sjtu-aithu system for dcase 2025 anomalous sound detection challenge," DCASE2025 Challenge, Tech. Rep., June 2025.

[8] A. Jiang, W. Liang, S. Feng, Y. Qiu, Y. Zhao, J. Li, P. Fan, W.-Q. Zhang, C. Lu, X. Chen, Y. Qian, and J. Liu, "Thuee system for dcase 2025 anomalous sound detection challenge," DCASE2025 Challenge, Tech. Rep., June 2025.

[9] B. Han, A. Jiang, X. Zheng, W.-Q. Zhang, J. Liu, P. Fan, and Y. Qian, "Exploring self-supervised audio models for generalized anomalous sound detection," *IEEE Transactions on Audio, Speech and Language Processing*, 2025.

[10] A. Jiang, X. Zheng, B. Han, Y. Qiu, P. Fan, W.-Q. Zhang, C. Lu, and J. Liu, "Adaptive prototype learning for anomalous sound detection with partially known attributes," in *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2025, pp. 1–5.

[11] 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," *In arXiv e-prints: 2606.01578*, 2026.

[12] 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 *Proceedings of the Detection and Classification of Acoustic Scenes and Events Workshop (DCASE)*, Barcelona, Spain, November 2021, pp. 1–5.

[13] 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 *Proceedings of the 7th Detection and Classification of Acoustic Scenes and Events 2022 Workshop (DCASE2022)*, Nancy, France, November 2022.

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

[15] S. Bharadwaj, S. Cornell, K. Choi, S. Fukayama, H.-j. Shim, S. Deshmukh, and S. Watanabe, "Openbeats: A fully open-source general-purpose audio encoder," in *2025 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. IEEE, 2025, pp. 1–5.

[16] W. Liang, Y. Qiu, A. Jiang, B. Han, T. Liu, X. Zheng, P. Fan, C. Lu, J. Liu, and W.-Q. Zhang, "Refgen: Reference-guided synthetic data generation for anomalous sound detection," in *ICASSP 2026-2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2026, pp. 15 877–15 881.

[17] B. Desplanques, J. Thienpondt, and K. Demuynck, "Ecapa-tdnn: Emphasized channel attention, propagation and aggregation in tdnn based speaker verification," *arXiv preprint arXiv:2005.07143*, 2020.

[18] K. Wilkinghoff, H. Yang, J. Ebbers, F. G. Germain, G. Wichern, and J. Le Roux, "Keeping the balance: Anomaly score calculation for domain generalization," in *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2025, pp. 1–5.

[19] M. Matsumoto, T. Fujimura, W. Huang, and T. Toda, "Adjusting bias in anomaly scores via variance minimization for domain-generalized discriminative anomalous sound detection," *Proc. DCASE*, 2025.

[20] S. Chen, M. Moradi, K. Paynabar, and H. Yan, "Flow mismatching: Unsupervised anomaly detection via velocity discrepancies in flow matching models," *arXiv preprint arXiv:2605.23070*, 2026.

[21] A. Jiang, W.-Q. Zhang, Y. Deng, P. Fan, and J. Liu, "Unsupervised anomaly detection and localization of machine audio: A gan-based approach," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023, pp. 1–5.