A TRAIN-NORMAL PROFILE ENSEMBLE FOR NOISE-AWARE UNSUPERVISED ANOMALOUS SOUND DETECTION

Technical Report

Hisataka Kajita *

Hisataka Kajita
Independent Researcher, Japan
tyomare@gmail.com

ABSTRACT

This report describes our submission to DCASE 2026 Challenge Task 2 [1], which targets first-shot unsupervised anomalous sound detection for machine condition monitoring under noise and domain shift. The submitted systems avoid using evaluation-test labels, source or target domain labels, and test attributes. Each evaluation machine is modeled only from its available normal training recordings. We build multiple interpretable anomaly-score components from robust acoustic statistics, two-channel near/far relations, percentile ranks, high-tail and low-tail departures, local-neighbor distances, and reconstruction-style residuals. Four submitted outputs differ only in their train-normal score aggregation rule: domain-conditioned consensus, top-tail rescue, two-sided rank scoring, and agreement-balanced consensus. Development experiments showed that simple single-model distances were insufficient, motivating multi-axis score families and conservative rank-shaping variants. For the final evaluation data, ground-truth labels are hidden, so official test performance is computed by the challenge organizers.

Index Terms-- anomalous sound detection, machine condition monitoring, unsupervised learning, score fusion, explainable audio analysis.

1. INTRODUCTION

DCASE 2026 Task 2 evaluates anomaly detection systems on machines that provide normal training recordings and unlabeled evaluation recordings. The evaluation set contains five machine types: BlowerDustCollector, Sander, SewingMachine, ToothBrush, and ToyDrone. The evaluation clips do not expose normal/anomaly labels, source/target domain labels, or test attributes. This setting discourages leaderboard-style threshold adjustment and requires a system that can infer abnormality from normal-machine structure alone [1].

The development and task formulation build on machine-sound datasets designed for domain-shift anomalous sound detection, including ToyADMOS2 and MIMII DG [2], [3]. Past DCASE Task 2 submissions indicate that no single representation is robust across all machine families. Strong systems often combine pretrained or handcrafted representations, normal-profile modeling, score normalization, and fusion. Our work follows that practical lesson, but keeps the final submitted system deliberately evaluation-safe: all submitted scores are derived from each machine's normal training audio and fixed aggregation rules.

The modeling principle is mechanism-oriented. We regard an anomaly as a break in mechanical-acoustic order, not as a label to memorize. We reasoned backward from plausible machine failure modes and implemented a bank of meta diagnostic axes: periodic order, spectral shape, temporal impacts and burstiness, component-level high/low energy, near/far channel consistency, and distributional departure from the train-normal profile. Each score component measures one axis before aggregation.

2. SUBMITTED SYSTEMS

The submitted systems share the same feature extraction and component-score bank. For every machine, train and test waveforms are converted into robust frame-level and clip-level acoustic descriptors. The descriptors are grouped by the diagnostic axes above: periodic-axis features measure autocorrelation and repeating structure; spectral-axis features measure spectral shape, flatness, contrast, and high-frequency residuals; temporal-axis features measure transients and burstiness; level/component-axis features measure broadband and band-limited high/low departures; and channel-axis features measure near/far level differences and spectral-gap descriptors.

For each descriptor group, robust medians and scales are fitted using only the normal training clips. Test clips are scored by several component functions: absolute robust z-score tails, top-k feature departures, Ledoit-Wolf Mahalanobis distance, PCA reconstruction residuals, k-nearest-neighbor distances, and empirical percentile ranks. Component scores are normalized against the training distribution before aggregation. The decision threshold is the 99th percentile of the corresponding train-normal score for the same machine and variant.

This differs from the official reconstruction-oriented AE/Mahalanobis baseline [4]: reconstruction residuals are retained only as one diagnostic axis inside a broader train-normal component bank.

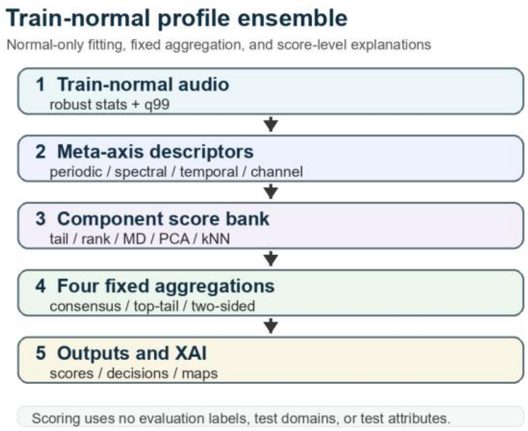

Fig. 1. Architecture of the submitted train-normal profile ensemble.

System	Aggregation rule	Purpose
Kajita_IND_task2_1	domain_consensus	Robust consensus; domain-aware when available.
Kajita_IND_task2_2	top_tail_rescue	Top-tail emphasis for sparse departures.
Kajita_IND_task2_3	two_sided_rank	High/low two-sided rank scoring.
Kajita_IND_task2_4	agreement_balanced	Agreement-regularized consensus.

3. DEVELOPMENT STUDY

For the report, we emphasize development validation under the same scoring restrictions used for the hidden evaluation submission: the system fits each machine from normal training clips only, does not use test labels, source/target test domains, or test attributes during scoring, and applies fixed aggregation and train-normal thresholds. Development labels are used only after scoring to compute AUC and pAUC. Earlier label-visible internal ablations were useful for choosing score families, but they are not reported here as expected performance of the four submitted systems.

System	Aggregation	Machine h-mean	Min machine
Kajita_IND_task2_1	domain_consensus	0.4966	0.4687
Kajita_IND_task2_2	top_tail_rescue	0.4875	0.3992
Kajita_IND_task2_3	two_sided_rank	0.4882	0.3932
Kajita_IND_task2_4	agreement_balanced	0.4877	0.3963

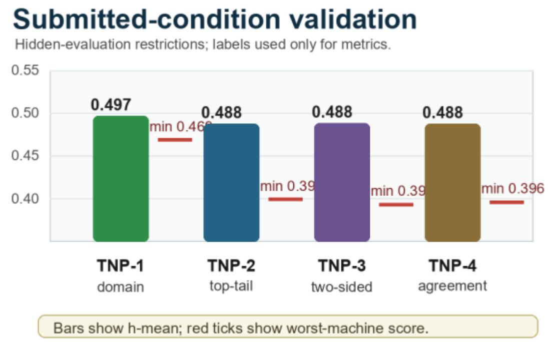

Fig. 2. Submitted-condition development validation for the four evaluation systems.

The four variants are therefore best understood as conservative rank-shaping alternatives over the same train-normal component bank, not as four independently trained model families. The small spread among their development scores is consistent with this design. The domain-consensus variant gives the best machine h-mean and the strongest worst-machine score in this submitted-condition replay, while the other variants provide alternative tail and agreement weighting for hidden evaluation.

4. EVALUATION DATA HANDLING

The official additional training and evaluation datasets were downloaded from Zenodo records 20151556 and 20437238. Each of the five evaluation machine folders contains 1000 normal training waveforms and 200 unlabeled test waveforms. Attribute files are available for the training side, but the submitted scoring path does not require test attributes.

Machine	Train clips	Test clips	Attributes
BlowerDustCollector	1000	200	yes
Sander	1000	200	yes
SewingMachine	1000	200	yes
ToothBrush	1000	200	yes
ToyDrone	1000	200	yes

The evaluation-test filenames follow the form `section_00_0000.wav`, so normal/anomaly and source/target labels cannot be inferred locally. We therefore report development results and leave official evaluation scoring to the challenge server.

5. EXPLAINABILITY

The system is designed to be explainable at the score-component level. For each machine and submitted variant, the pipeline writes component-score CSV files and profile JSON summaries. Additional diagnostic figures show train-normal and test score distributions, component maps, and the mean positive contribution of the most active feature groups. These artifacts allow us to inspect whether a clip is scored highly because of high-side bursts, low-side missing energy, periodic instability, spectral-shape departure, or near/far channel disagreement.

This explanation style is intentionally tied to the submitted score rather than to a separate post-hoc classifier. It is useful for finding failure modes such as domain-like normal variation being mistaken for an anomaly, or a single unstable feature group dominating the final rank.

6. DISCUSSION

The main tradeoff in this submission is between development-set exploration and evaluation-set portability. Label-visible internal experiments helped us identify useful score families, but the final evaluation hides the labels and test domains and uses new machine types. The submitted systems therefore emphasize reproducible normal-only modeling and avoid any test-distribution retuning. We expect this to be conservative, but it is aligned with the task constraints and gives robust outputs for all official machines.

7. CONCLUSION

We submitted four train-normal profile ensemble systems for DCASE 2026 Task 2. The systems use robust acoustic descriptors, two-channel relation features, component-level anomaly scoring, and fixed aggregation rules. Submitted-condition development validation shows modest but interpretable differences among the four rank-shaping variants over the same normal-profile component bank. The final submitted systems are conservative unknown-machine variants that can be run without evaluation labels, test domains, or test attributes.

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," In arXiv e-prints: 2606.01578, 2026.

[2] 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," Proceedings of the Detection and Classification of Acoustic Scenes and Events Workshop (DCASE), pp. 1-5, Barcelona, Spain, 2021.

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

[4] 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.