# ADAPTIVE FRAMEWORK FOR FIRST-SHOT UNSUPERVISED ANOMALOUS SOUND DETECTION IN MACHINE MONITORING

Technical Report

*Jie Yang*

Shanghai, China
1227897389@qq.com

**ABSTRACT**

This technical report details our approach to addressing Task 2 of the DCASE 2024 Challenge, which centers on First-Shot Unsupervised Anomalous Sound Detection for Machine Condition Monitoring. The objective of this year's challenge is to create a system that functions efficiently regardless of the presence of attribute information, mirroring real-world situations where such data may be intermittently accessible.

To tackle this challenge, we propose an adaptive anomaly detection framework that adjusts seamlessly to the availability of attribute information. Our method employs an attribute classification scheme for detecting anomalous sounds. In cases where attribute information is present, we perform detailed anomaly detection by subdividing all attributes. For situations lacking attribute information, we utilize domain-specific information for effective detection.

The adaptive system achieved a harmonic mean of 57.75% across all machine types and domains for both AUC and pAUC (p= 0.1) on the development set. This result demonstrates significant improvement and ensures the system's adaptability under varying data conditions. Moreover, the framework's flexibility in handling different types of input data enhances its applicability in real-world industrial machine monitoring scenarios.

***Index Terms*—** Anomalous Sound Detection, Machine Condition Monitoring, Real-World Scenarios, Adaptive Anomaly Detection Framework, Attribute Classification Scheme

## 1. INTRODUCTION

With the advancements in society and technology, machinery has become increasingly crucial in industrial production. However, various factors often lead to equipment failure during operation, which can impact production efficiency and performance, and even result in severe safety incidents. Anomalous sound detection (ASD) is a vital component of machine condition monitoring tools that involves identifying whether the sound emitted from a target machine is normal or anomalous [1].

The task topic of this year's task2 is First-shot problem under attribute-available and unavailable conditions [2], which is the follow-up task from DCASE 2020 task2 to DCASE 2023 task2 [3].

In view of the above situation, we propose an adaptive anomaly detection framework, which can ensure that the framework can operate efficiently regardless of the existence of attribute information, and is closer to the actual situation in industry. We mainly adopt an attribute classification scheme to model uniformly for all machines. For machines with attributes, we perform detailed anomaly detection by subdividing by all attributes. In the absence of attribute information, we utilize domain-specific information for effective detection. Finally, we extracted the embedding vectors through MobileFaceNet [4], and used arcface [5] as the loss function of the model, and the back-end used an anomaly detection algorithm to evaluate the degree of abnormality.

## 2. METHODOLOGY

### 2.1. Classification model

The entire abnormal sound detection system comprises a front-end feature extractor and a back-end anomaly detector. The front-end feature extractor serves as a supervised attribute classification model, which is constructed on the MobileFaceNet network structure. This attribute classification model is simultaneously supervised by machine type, attribute, and domain information. To minimize intra-class distance and maximize inter-class distance, ArcFace is employed as the model's loss function. The Kmeans algorithm determines the class center, and the anomaly score is evaluated based on the minimum distance between the test sample and each class center, thereby facilitating back-end anomaly detection.

### 2.2. Feature extraction

In the step of feature extraction, we deeply analyze four types of feature images, including spectrogram, original waveform graph, Mel-spectrogram and Short-Time Fourier Transform (STFT) spectrogram. In order to ensure the validity and accuracy of the selected features, we performed combined validation of these features from multiple perspectives on the development set. After rigorous experiments and analysis, two feature images, Mel-spectrogram and STFT spectrogram, are finally selected as input. Meanwhile, at the physical level, these two features focus on sound and vibration information, respectively. For both feature representations, we used the magnitude spectrogram with window length 1024 and window shift size 512. These two kinds of features are respectively fed into the embedding vector extraction network based on MobileFaceNet for processing. Through this process, we obtain two kinds of feature vectors and combine them to form the final embedding feature vector.

### 2.3. Back-end anomaly detector

The back-end anomaly detector of the system consists of three steps [6]. For the source domain, the k-means algorithm is applied to obtain multiple class centers for each machine type, and then all cosine distances between the given test sample and these class centers are calculated. While for the target domain, the cosine distance between the given test sample and all 10 normal samples of the same machine type is calculated. Finally, for a test sample, the minimum value among all cosine distances calculated for this sample is taken as the anomaly score for this sample. Thus, the anomaly score of the sample can indicate whether the sample is abnormal.

## 3. RESULTS

The proposed system is compared with the benchmark system of DCASE 2024 in Challenge Task 2, namely Baseline MSE and Baseline MAHALA [7]. Our system outperforms the baseline system for most of the machines in the development set, as shown in Table 1.

Table 1: Anomaly detection results for the proposed system

| | Method | Baseline MSE | Baseline MAHALA | Proposed system |
|---|---|---|---|---|
| ToyCar | AUC(source) | 66.98% | 63.01% | 54.39% |
| | AUC(target) | 33.75% | 37.35% | 56.87% |
| | pAUC | 48.77% | 51.04% | 47.28% |
| ToyTrain | AUC(source) | 76.63% | 61.99% | 57.36% |
| | AUC(target) | 46.92% | 39.99% | 53.59% |
| | pAUC | 47.95% | 48.21% | 47.79% |
| bearing | AUC(source) | 62.01% | 54.43% | 51.26% |
| | AUC(target) | 61.4% | 51.58% | 54.71% |
| | pAUC | 57.58% | 58.82% | 57.09% |
| fan | AUC(source) | 67.71% | 79.37% | 62.03% |
| | AUC(target) | 55.24% | 42.7% | 66.27% |
| | pAUC | 57.53% | 53.44% | 54.98% |
| gearbox | AUC(source) | 70.4% | 81.82% | 77.46% |
| | AUC(target) | 69.34% | 74.35% | 75.20% |
| | pAUC | 55.65% | 55.74% | 60.04% |
| slider | AUC(source) | 66.51% | 75.35% | 74.24% |
| | AUC(target) | 56.01% | 68.11% | 76.38% |
| | pAUC | 51.77% | 49.05% | 55.65% |
| valve | AUC(source) | 51.07% | 55.69% | 51.37% |
| | AUC(target) | 46.25% | 53.61% | 55.86% |
| | pAUC | 52.42% | 51.26% | 49.98% |
| All (hmean) | AUC(source) | 65.00% | 65.77% | 59.68% |
| | AUC(target) | 50.28% | 49.51% | 61.46% |
| | pAUC | 52.84% | 52.28% | 52.86% |

## 4. CONCLUSION

In this technical report, we introduce our submission systems to DCASE 2024 challenge task 2. We propose an attribute classification scheme based on MobileFaceNet network, combined with the kmeans algorithm of the back-end, to achieve accurate detection of abnormal sounds. When the attribute information is complete, we make a detailed division of each type of attribute, so as to realize the depth monitoring of abnormal conditions. Even in the case of missing attribute information, we divide the domain information as attributes to effectively identify abnormal conditions.

In addition, in the feature extraction stage, we comprehensively consider the characteristics of the machine in the two key dimensions of sound and vibration in the industrial environment. We combine the advantages of Mel-spectrogram and Short-Time Fourier Transform (STFT) spectrogram, and this innovative initiative significantly improves the universality and effectiveness of anomaly detection systems. This integrated feature extraction approach ensures that the system can more accurately identify and process abnormal sound data from different sources and environments.

## 5. REFERENCES


[1] Nishida, Tomoya and Harada, Noboru and Niizumi, Daisuke and Albertini, Davide and Sannino, Roberto and Pradolini, Simone and Augusti, Filippo and Imoto, Keisuke and Dohi, Kota and Purohit, Harsh and Endo, Takashi and Kawaguchi, Yohei. Description and Discussion on DCASE 2024 Challenge Task 2: First-Shot Unsupervised Anomalous Sound Detection for Machine Condition Monitoring. In arXiv e-prints: 2406.07250, 2024.

[2] Harada, Noboru and Niizumi, Daisuke and Takeuchi, Daiki and Ohishi, Yasunori and Yasuda, Masahiro and Saito, Shoichiro. 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), 1–5. Barcelona, Spain, November 2021.

[3] Dohi, Kota and Nishida, Tomoya and Purohit, Harsh and Tanabe, Ryo and Endo, Takashi and Yamamoto, Masaaki and Nikaido, Yuki and Kawaguchi, Yohei. 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.

[4] S. Chen, Y. Liu, X. Gao, and Z. Han, "Mobilefacenets: Efficient cnns for accurate real-time face verification on mobile devices," in Chinese Conference on Biometric Recognition. Springer, 2018, pp. 428–438.

[5] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019, pp. 4690–4699.

[6] Wilkinghoff K. Design Choices for Learning Embeddings from Auxiliary Tasks for Domain Generalization in Anomalous Sound Detection[C]//ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2023: 1-5.

[7] Harada, Noboru and Niizumi, Daisuke and Ohishi, Yasunori and Takeuchi, Daiki and Yasuda, Masahiro. First-Shot Anomaly Sound Detection for Machine Condition Monitoring: A Domain Generalization Baseline. In 2023 31st European Signal Processing Conference (EUSIPCO), 191-195. 2023.