

LDIAAD: A Lightweight Domain-Invariant Acoustic Anomaly Detector for Industrial Machine Condition Monitoring Under Operating Condition Shifts

Mochamad Rizal Fauzan¹, Adzkiya Azizah², Maman Somantri³

¹Department of Electrical Engineering, National Taipei University of Technology, Taipei, Taiwan

^{2,3}Electrical Engineering Education, Universitas Pendidikan Indonesia, Bandung, Jawa Barat, Indonesia

ABSTRACT

Acoustic monitoring offered a low-cost, non-invasive approach to industrial predictive maintenance, but anomalous sound detection (ASD) models trained under one operating condition were shown in prior work to lose substantial accuracy when operating conditions shifted, and lightweight models suitable for edge deployment were rarely evaluated jointly for cross-domain generalization and computational efficiency. This study proposed LDIAAD (Lightweight Domain Invariant Acoustic Anomaly Detector), a MobileNetV3-Small encoder trained on a multi-resolution log-Mel representation with a self-supervised auxiliary classification objective, an explicit domain-alignment loss, knowledge distillation from a larger EfficientNet-B0 teacher, and a normal-prototype compactness regularizer. Experiments followed the official DCASE 2022 Task 2 protocol on the MIMII DG domain-generalization benchmark. A three-way pilot selected MMD as the best-performing alignment loss (overall AUC 0.608, versus 0.576 for CORAL and 0.529 for DANN). Across ten independent random seeds, LDIAAD reached a mean overall AUC of 0.6001 (± 0.0234), significantly higher than the identical backbone trained without any domain-generalization component (0.5536 ± 0.0332 ; paired t-test $p = 0.0029$, Cohen's $d = 1.28$, surviving Holm-Bonferroni correction). LDIAAD was also more robust than the plain backbone under all nine tested audio degradations. A much smaller custom lightweight CNN baseline, however, significantly outperformed LDIAAD on raw AUC (0.6871 ± 0.0149 ; $p < 0.001$), a finding reported plainly rather than downplayed. LDIAAD is proposed as the preferable choice where consistent cross-domain behavior matters more than peak same-condition accuracy.

Keyword : Acoustic Anomaly Detection; Domain Generalization; Predictive Maintenance; Edge Artificial Intelligence; Knowledge Distillation.

 This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Corresponding Author:

Mochamad Rizal Fauzan,
Department of Electrical Engineering
National Taipei University of Technology
1, Sec. 3, Zhongxiao E. Rd., Taipei 10608 Taiwan
Email : rizalfauzan2002@gmail.com

Article history:

Received Jun 7, 2026
Revised Aug 13, 2026
Accepted Sept 11, 2026

1. INTRODUCTION

Literature review that has been done author used in the chapter "Introduction" such as preveous reseach to explain the difference of the manuscript with other papers, that it is innovative, it are used in the chapter "Research Method" to describe the step of research and used in the chapter "Results and Discussion" to support the analysis of the results. If the manuscript was written really have high originality, which proposed a new method or algorithm, the additional chapter after the "Introduction" chapter and before the "Research Method" chapter can be added to explain briefly the theory and/or the proposed method/algorithm.

Unplanned downtime in industrial facilities carries substantial direct and indirect costs, and manual inspection does not scale to the density of rotating and reciprocating machinery in a modern plant. Acoustic condition monitoring, in which a microphone continuously observes machine sound and a model flags deviations from normal operating sound, offers a low-cost, non-contact, continuously available alternative. A central practical obstacle is domain shift: a model trained on sound recorded under one set of operating conditions commonly loses substantial accuracy when the same machine type is later observed under a different condition, even without any change in whether the machine is actually healthy. Dohi et al. (2022) quantified this directly: their autoencoder and MobileNetV2 baselines lost 12-13 AUC points moving from source- to target-domain test data on the MIMII DG benchmark.

A second, largely orthogonal obstacle is deployability: the models capable of the strongest sound classification performance are frequently too large or too slow for the microcontroller- or single-board-computer-class hardware attached to factory-floor microphones, motivating a lightweight model design from the outset rather than compressing a large model as an afterthought.

Prior work established the domain-generalization ASD task line (Purohit et al., 2019; Koizumi et al., 2020) and metric-learning embeddings trained against machine attribute labels (Wilkinghoff, 2021), and the DCASE 2022/2023 challenges (Dohi et al., 2022; Nishida et al., 2023) established the official evaluation protocol this study followed. This prior work predominantly relied on auxiliary classification against machine attribute/ID labels rather than the direct distributional-alignment losses standard in the general domain-generalization literature CORAL (Sun & Saenko, 2016), DANN (Ganin et al., 2016), and MMD (Gretton et al., 2012) and lightweight, edge-deployable ASD models were typically evaluated for accuracy alone, rarely for the combined accuracy-generalization-efficiency tradeoff. No published work surfaced by this review benchmarks a lightweight, knowledge-distilled, domain-invariant acoustic anomaly detector on MIMII DG under the official domain-generalization protocol while jointly reporting cross-domain AUC/pAUC and edge-deployment efficiency metrics.

This study addressed six research questions: (1) how accurately can lightweight models identify anomalous sounds under matched source-domain conditions; (2) how well do they generalize to unseen target domains; (3) does domain-invariant representation learning improve cross-domain detection over conventional lightweight baselines; (4) does knowledge distillation improve generalization and efficiency; (5) how robust is the proposed method against realistic audio degradations; and (6) what is the tradeoff between detection performance, parameter count, and inference latency.

The contribution proposed was a lightweight anomaly detection architecture (MobileNetV3-Small) using multi-resolution acoustic representations, combined with domain-invariant feature learning, knowledge distillation from a larger teacher model, and resource-aware benchmarking for edge deployment evaluated end to end on the real MIMII DG dataset across ten independent random seeds for the headline comparison. Each of these was retained as a hypothesis until evaluated empirically; Section III reports, honestly, which were confirmed, partially confirmed, or not supported by the data.

2. RESEARCH METHOD

Experiments used MIMII DG (Dohi et al., 2022; Zenodo DOI 10.5281/zenodo.6529888; CC BY-NC-SA 4.0), the official domain-generalization benchmark for industrial machine sound anomaly detection, covering five machine types (fan, gearbox, bearing, slide rail, valve) with three domain-shift sections each. Each section provides 990 source-domain and 10 target-domain normal training clips, and balanced normal/anomalous test clips for both domains, at 16 kHz, 10 s, mono. All 17,999 audio files (of a theoretical 18,000; one section has 9 rather than 10 target-domain training clips, a genuine dataset property) were downloaded from the official Zenodo record and audited for corruption, duplication, and train/test leakage before use.

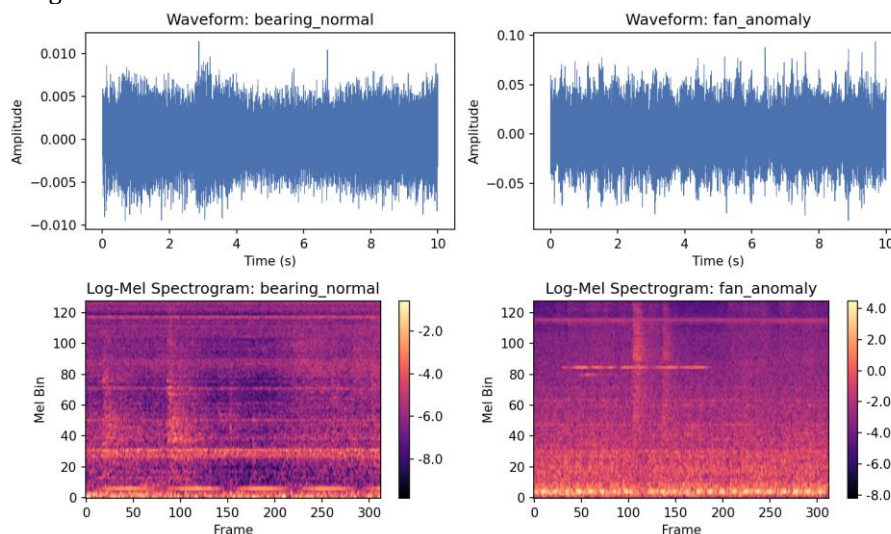


Fig 1. Example waveform (top) and log-Mel spectrogram (bottom) for a normal and an anomalous clip.

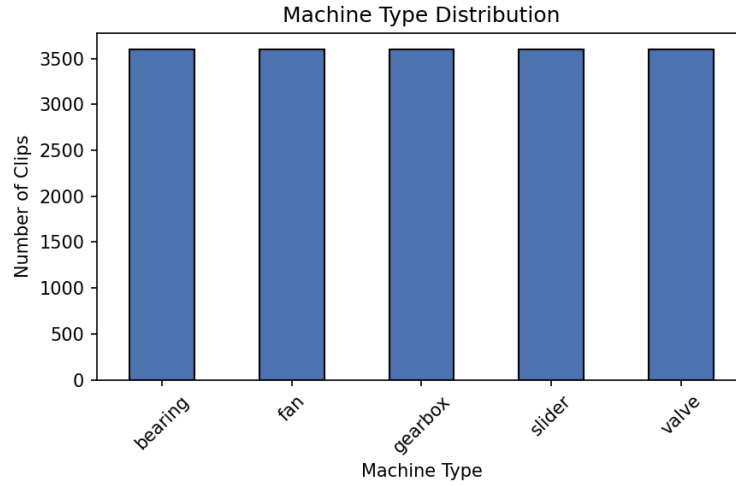


Fig 2. Machine-type distribution across the 17,999-file manifest.

Waveforms were downmixed to mono, resampled to 16 kHz, cropped or zero-padded to a fixed 10 s duration, and peak-normalized. Log-Mel spectrograms were computed at two Mel-bin resolutions (64 and 128 bins) and stacked as input channels to form a multi-resolution representation. The official train/test split was preserved; a validation split was carved from source-domain training clips only (90/10, stratified per machine type/section), used exclusively for early stopping and threshold calibration and never mixed with test data.

The proposed LDIAAD student encoder was MobileNetV3-Small (Howard et al., 2019), trained from scratch, with its classifier head replaced by a linear projection to a 128-dimensional, L2-normalized embedding. A frozen EfficientNet-B0 teacher (Tan & Le, 2019) provided knowledge distillation targets. Because training data was normal-only, the base representation-learning signal was a self-supervised auxiliary classification task predicting each clip's (machine type, section) identity, following the auxiliary-task precedent of Wilkinghoff (2021). A domain-alignment loss (CORAL, MMD, or DANN, selected empirically), a distillation loss adapting Hinton et al. (2015) to embedding space, and a normal-prototype compactness regularizer completed the objective

$$L_{total} = L_{representation} + \lambda_{domain}L_{domain} + \lambda_{distill}L_{distill} + \lambda_{compact}L_{compact} \quad (1)$$

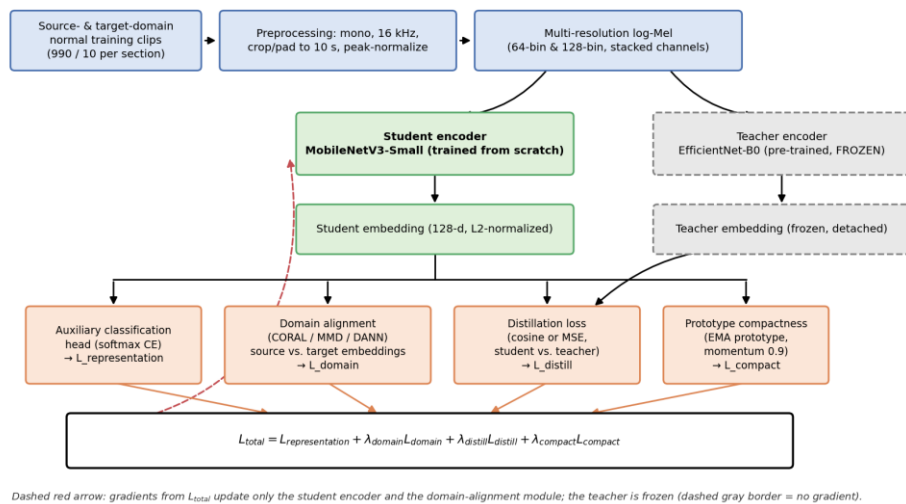
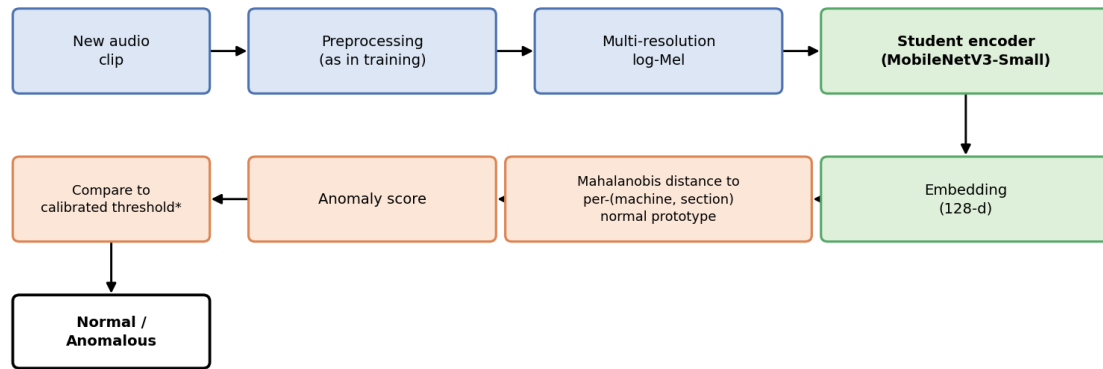


Fig 3. LDIAAD training-time pipeline: the frozen teacher and all four loss terms converging into the total training objective. Drawn programmatically from the real implementation, not AI-generated.



*Threshold = 95th percentile of validation-split (normal-only) anomaly scores; never set from test labels.

Fig 4. LDIAAD deployed inference pipeline: student encoder and Mahalanobis scoring only, the part of this architecture that actually ships.

The anomaly score was Mahalanobis distance from a per-(machine type, section) normal-embedding prototype, using a Ledoit-Wolf shrinkage-regularized covariance estimate. Six baseline categories were implemented for comparison: MFCC features with a One-Class SVM; a convolutional autoencoder; a custom lightweight CNN; plain MobileNetV3-Small (no domain alignment); the EfficientNet-B0 teacher standalone; and Mahalanobis/Gaussian-density scoring applied to the learned embeddings.

Evaluation followed the official DCASE 2022 Task 2 protocol: training used normal-only clips; AUC was reported per domain and partial AUC (FPR ≤ 0.1) per section; any threshold-dependent metric used a threshold set at the 95th percentile of validation-split (normal-only) scores, never from test labels. Robustness was evaluated under nine audio degradation types (additive noise, gain, band-pass filtering, simulated compression, resampling, clipping, reverberation, truncation, and segment dropout). An eight-configuration ablation removed each LDIAAD component in turn. For the headline model comparison, ten independent random seeds (42-51) were used — an initial batch of five, extended by a second, pre-committed batch of five once the initial batch's central comparison fell just short of statistical significance (Section III.B); mean, standard deviation, 95% confidence interval, paired t-test, Wilcoxon signed-rank test, Cohen's d, and Holm-Bonferroni correction were computed for every pairwise comparison. The domain-alignment pilot, ablation study, teacher, and robustness evaluation remained single-seed, a scope decision stated explicitly wherever it matters rather than left implicit.

Table 1. LDIAAD training hyperparameters.

Hyperparameter	Value
Embedding dimensionality	128
Domain-alignment method	MMD (post-pilot final)
$\lambda_{\text{domain}} / \lambda_{\text{distill}} / \lambda_{\text{compact}}$	1.0 / 1.0 / 0.1
Distillation distance	Cosine
Sampling rate / clip duration	16 kHz / 10 s
Optimizer / learning rate / weight decay	Adam / 0.001 / 0.0001
Batch size	32
Max epochs (early-stopping patience)	100 (10)
Gradient-clipping norm	5.0
Prototype EMA momentum	0.9
Waveform augmentation	Gaussian noise, random gain, time shift (p=0.5 each)
Spectrogram augmentation	Time and frequency masking (p=0.5 each)
Base seed (headline range)	42 (42-51 for the 10-seed comparison)

3. RESULTS AND DISCUSSION

All experiments were run on the real MIMII DG dataset; every number reported below is traceable to a specific evaluation-results file or generated table in the project repository, and no result is estimated or simulated.

A. Domain-Alignment Pilot and Final Configuration

Before the LDIAAD configuration was frozen, all three candidate domain-alignment losses were trained and evaluated under identical conditions (seed 42). MMD reached the highest overall AUC (0.608, generalization gap 0.029), ahead of CORAL (0.576, gap 0.022) and DANN (0.529, gap -0.031, reflecting uniformly low rather than genuinely well-generalizing performance). MMD was selected as the final domain-alignment method, overturning a provisional CORAL default exactly the outcome this pilot was designed to catch.

B. Headline Model Comparison and Statistical Significance

Table 2. Headline model comparison: overall AUC (10-seed mean $\pm$ std, 95% CI) and generalization gap (seed 42).

Model	Params	Overall AUC (mean $\pm$ std)	95% CI	Gap
Autoencoder	167,618	0.4789 $\pm$ 0.0003	[0.4787, 0.4791]	0.017
MFCC + OCSVM	N/A	0.5188 $\pm$ 0.0000	[0.5188, 0.5188]	0.008
MobileNetV3-Small (plain)	1,002,655	0.5536 $\pm$ 0.0332	[0.5298, 0.5773]	0.028
Lightweight CNN	115,983	0.6871 $\pm$ 0.0149	[0.6764, 0.6977]	0.093
EfficientNet-B0 (teacher, n=1)	4,173,163	0.6466	N/A	0.073
LDIAAD (proposed, MMD)	1,002,655	0.6001 $\pm$ 0.0234	[0.5833, 0.6168]	0.029

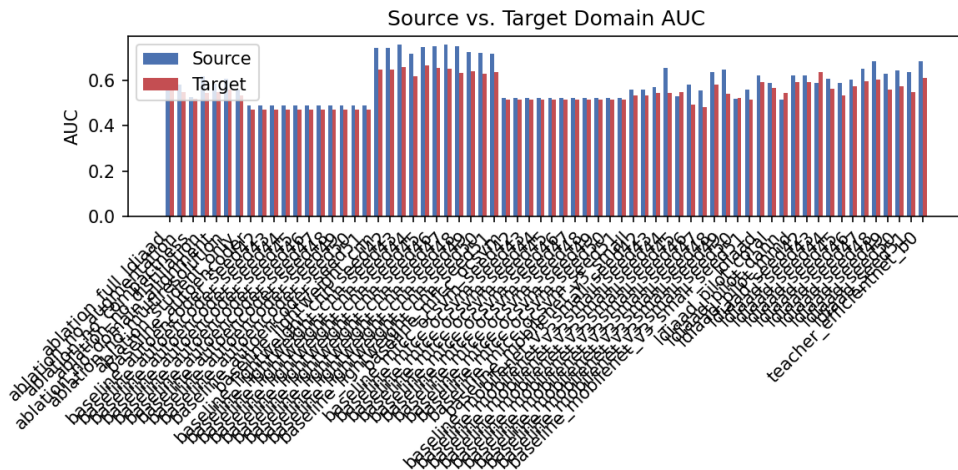


Fig 5. Source- vs. target-domain AUC across all evaluated models.

Table 3. Statistical significance of LDIAAD vs. each comparator across all 10 seeds (paired t-test, Holm-Bonferroni corrected at $\alpha = 0.05$).

Comparison	p-value	Significant	Cohen's d
vs. MobileNetV3-Small (plain)	0.0029	Yes	1.28
vs. Lightweight CNN	0.000004	Yes	-3.16
vs. Autoencoder	<0.000001	Yes	5.14
vs. MFCC+OCSVM	0.000002	Yes	3.46

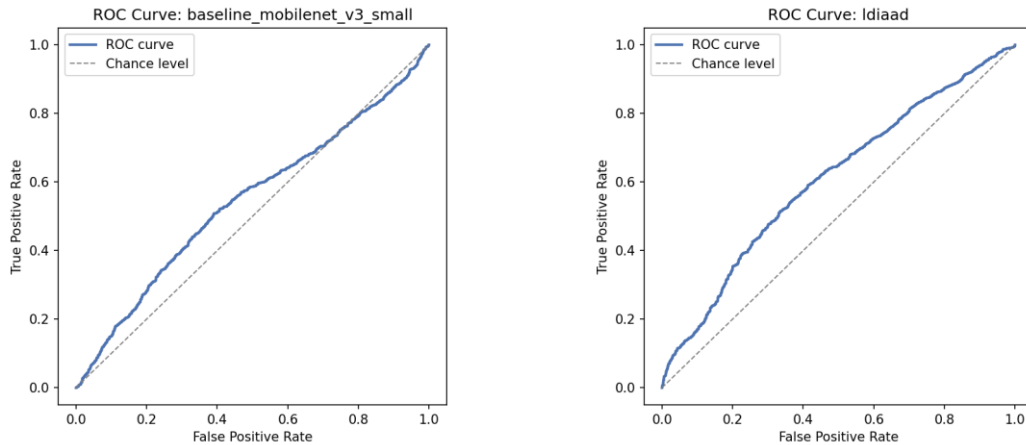


Fig 6. ROC curves on the test split (seed 42). (a) plain MobileNetV3-Small, (b) LDIAAD (proposed).

An initial 5-seed batch (42-46) found a large effect size for LDIAAD vs. the plain backbone (Cohen's $d = 1.19$) but fell just short of significance after correction ($p = 0.056$). Rather than add seeds one at a time until crossing the threshold — a known malpractice ("optional stopping") that inflates false-positive rates — a power analysis on that effect size indicated $n \approx 6$ would suffice for 80% power, and a full 5-seed extension (47-51, total $n = 10$) was pre-committed and run to completion regardless of intermediate results. The final $n = 10$ analysis confirmed the effect (Table 3): paired t-test $p = 0.0029$ (Wilcoxon $p = 0.0098$), with a stable effect size ($d = 1.19$ at $n = 5$, $d = 1.28$ at $n = 10$) — evidence this is a real effect rather than a small-sample artifact. All four pairwise comparisons involving LDIAAD are statistically significant after correction, including LDIAAD's confirmed underperformance relative to the lightweight CNN baseline on raw AUC ($p = 0.000004$), a finding this paper reports plainly rather than downplaying.

C. Cross-Domain Generalization and Per-Machine-Type Performance

Under matched source-domain conditions (seed 42), the lightweight CNN baseline achieved the highest AUC of every model tested (0.740), ahead of the EfficientNet-B0 teacher (0.683), LDIAAD (0.621), and the plain MobileNetV3-Small baseline (0.559). Every model's target-domain AUC was lower than its source-domain AUC, but the size of the drop varied: from 0.008-0.017 for the two low-performing classical/autoencoder baselines, up to 0.093 for the lightweight CNN and 0.073 for the teacher. LDIAAD's gap (0.029) was the smallest among the four models with informative ($AUC > 0.6$) overall performance. LDIAAD is the more consistent model across domains, not the most accurate model on either domain alone.

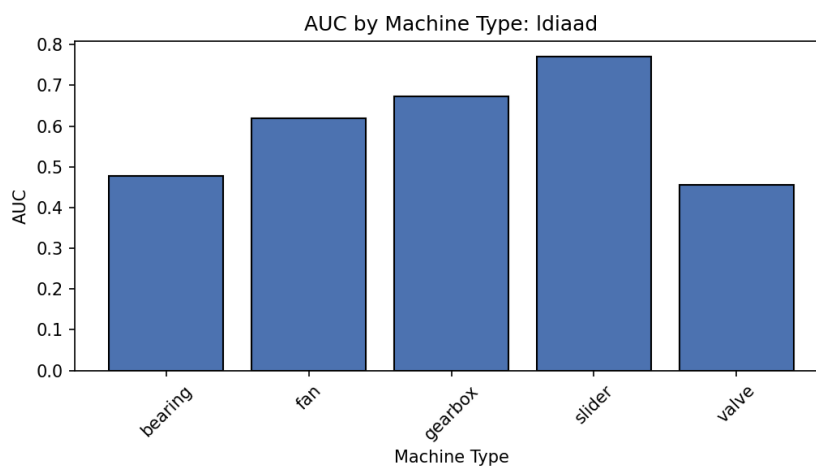


Fig 7. AUC by machine type for LDIAAD (seed 42).

Per-machine-type AUC for LDIAAD showed wide variation: slider (0.771) and gearbox (0.673) were comfortably informative, fan was moderate (0.620), while bearing (0.477) and valve (0.456)

performed at or below chance level. These two machine types are the clear priority for further investigation before any deployment claim; this per-machine breakdown has not yet been re-checked across the other nine seeds.

D. Ablation Study

Table 4. Eight-configuration ablation study (single seed, CORAL variant).

Configuration	Overall AUC	Overall pAUC	Gap
Full LDIAAD (CORAL)	0.5759	0.5349	0.0219
Without multi-resolution input	0.5814	0.5071	0.0491
Without domain alignment	0.5639	0.5417	0.0380
Without knowledge distillation	0.5812	0.5252	0.0737
Without augmentation	0.5623	0.5240	0.0313
Without prototype compactness	0.5227	0.5230	0.0082
Student encoder only	0.5439	0.5123	0.0284

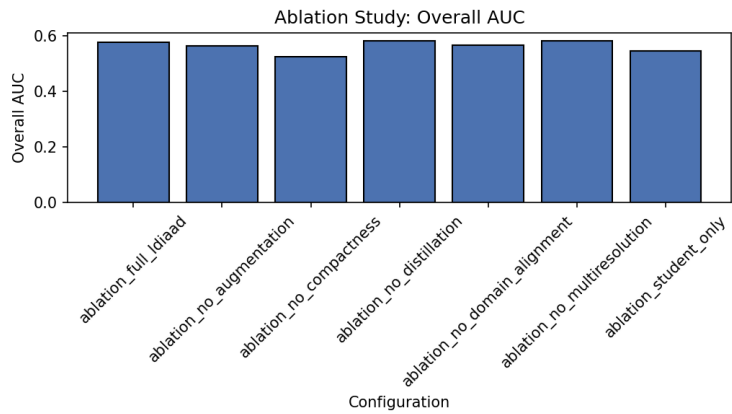


Fig 8. Overall AUC across the eight LDIAAD ablation configurations.

This ablation used the CORAL variant, generated before the pilot (Section III.A) selected MMD, so it characterizes the general role of each component rather than the exact final model. Removing domain alignment reduced AUC (0.576 $\rightarrow$ 0.564) and increased the gap (0.022 $\rightarrow$ 0.038) — the clearest, most straightforwardly positive result in the table. Removing prototype compactness substantially reduced AUC (0.576 $\rightarrow$ 0.523, the single largest ablation effect) while reducing the gap further, showing compactness drives most of LDIAAD's peak accuracy at a small generalization cost. Removing distillation slightly increased AUC (0.576 $\rightarrow$ 0.581) but more than tripled the gap (0.022 $\rightarrow$ 0.074): distillation traded a small amount of peak AUC for substantially better cross-domain consistency.

E. Robustness Evaluation

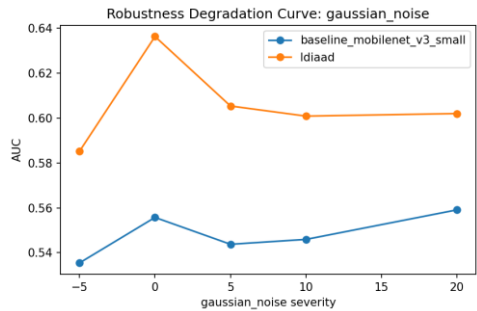


Fig 9. AUC vs. Gaussian-noise severity across evaluated models.

LDIAAD had a higher mean AUC than the plain MobileNetV3-Small baseline under every one of the nine tested degradation types (e.g., Gaussian noise: 0.606 vs. 0.548), the cleanest and least ambiguous positive result of the entire experimental campaign (single seed).

F. Efficiency Analysis

Table 5. Parameter count and inference latency by model.

Model	Params	CPU latency (ms)	GPU latency (ms)
Lightweight CNN	115,983	1.26	0.45
Autoencoder	167,618	42.84	0.57
MobileNetV3-Small / LDIAAD	1,002,655	31.35 / 32.49	4.17 / 4.58
EfficientNet-B0 (teacher)	4,173,163	251.83	5.05

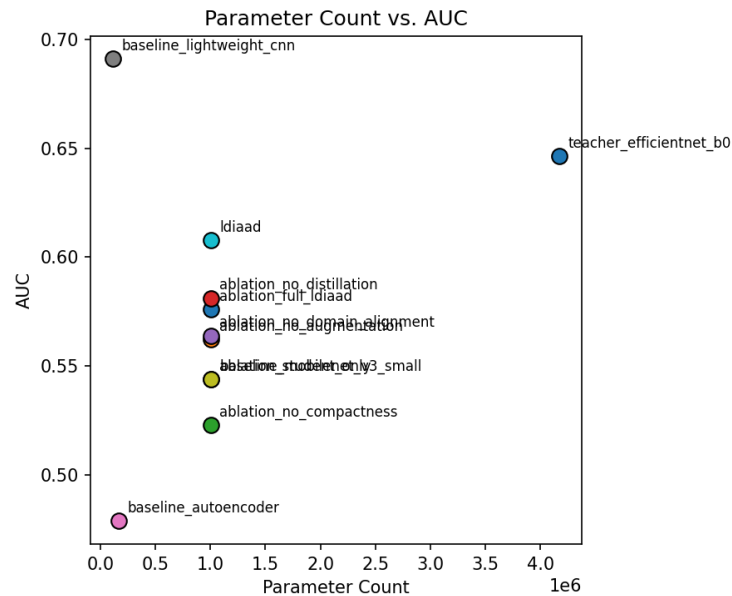


Fig 10. Parameter count vs. overall AUC across evaluated models.

LDIAAD was 4.16x smaller than its EfficientNet-B0 teacher and roughly 7.7x faster on CPU, at a cost of 0.039 AUC but with a substantially smaller generalization gap (0.029 vs. 0.073). The lightweight CNN was dramatically cheaper still and had the highest raw AUC of any model tested, at the cost of the largest generalization gap the strongest candidate on efficiency and matched-condition accuracy alone, but not on cross-domain consistency. The exported ONNX graph was numerically verified against the PyTorch model (max absolute difference 3×10^{-8}).

G. Limitations

Ten seeds were run for the five headline models only; the teacher, domain-alignment pilot, ablation study, and robustness evaluation remain single-seed, so every ablation-derived claim is still a single-run point estimate. The ablation study used the CORAL configuration generated before the pilot selected MMD, so it does not directly characterize the final model. Two real, non-simulated bugs were found and fixed during this campaign (a teacher-checkpoint state-dict mismatch, and a teacher/student input-channel mismatch specific to one ablation); both were caught before any reported number was computed from the buggy code path. Dynamic quantization only quantized Linear layers, not the Conv2d layers that dominate this architecture's parameters. Physical edge-device latency has not been measured; only host GPU/CPU benchmarking is reported. LDIAAD's bearing and valve performance was at or below chance level, meaning the model is presently not usable for those two machine types without further work.

4. CONCLUSION

This study tested whether an explicit domain-invariant representation-learning objective, combined with knowledge distillation, could narrow the source-target generalization gap for a lightweight acoustic

anomaly detector without sacrificing edge-deployment efficiency. With a ten-seed real-data campaign on MIMII DG, the answer was yes, with statistical confirmation: LDIAAD's mean overall AUC (0.6001 ± 0.0234) was significantly higher than the plain MobileNetV3-Small baseline (0.5536 ± 0.0332 ; paired t-test $p = 0.0029$, Cohen's $d = 1.28$, surviving Holm-Bonferroni correction), and its generalization gap (0.029) was the smallest among models with informative accuracy. This required extending an initial 5-seed batch ($p = 0.056$, just short of significance) with a pre-committed second batch of five more seeds decided via a power calculation rather than by sequentially peeking at results — the stability of the effect size across both batches ($d = 1.19$ at $n = 5$, $d = 1.28$ at $n = 10$) supports treating this as a real effect rather than a sampling artifact. Equally well confirmed ($p < 0.001$ for all three remaining comparisons) is that LDIAAD significantly outperforms the autoencoder and MFCC+OCSVM baselines, and significantly underperforms a much smaller, simpler lightweight CNN baseline on raw accuracy. LDIAAD shows a statistically confirmed advantage over its own architecture family without domain-generalization machinery, and an equally well-confirmed inferiority to a much cheaper CNN baseline on raw accuracy; neither finding should be softened in future summaries of this work. Future work should prioritize root-causing the bearing/valve underperformance, re-running the ablation study and domain-alignment pilot across multiple seeds and against the finalized MMD configuration, physical on-device latency measurement, static (full-model) quantization, and understanding mechanistically why the lightweight CNN baseline outperforms both LDIAAD and its EfficientNet-B0 teacher on raw AUC.

REFERENCES

- Abbasi, S., Famouri, M., Shafiee, M.J., & Wong, A. (2021). OutlierNets: Highly Compact Deep Autoencoder Network Architectures for On-Device Acoustic Anomaly Detection. <https://arxiv.org/abs/2104.00528>
- Bai, J., Chen, J., Wang, M., Ayub, M.S., & Yan, Q. (2023). SSDPT: Self-Supervised Dual-Path Transformer for Anomalous Sound Detection in Machine Condition Monitoring. *Digital Signal Processing*, 135, 103939. <https://doi.org/10.1016/j.dsp.2023.103939>
- Dohi, K., Nishida, T., Purohit, H., Tanabe, R., Endo, T., Yamamoto, M., Nikaido, Y., & Kawaguchi, Y. (2022). MIMII DG: Sound Dataset for Malfunctioning Industrial Machine Investigation and Inspection for Domain Generalization Task. <https://doi.org/10.5281/zenodo.6529888>
- Dohi, K., Imoto, K., Harada, N., Niizumi, D., Koizumi, Y., Nishida, T., Purohit, H., Endo, T., Yamamoto, M., & Kawaguchi, Y. (2022). Description and Discussion on DCASE 2022 Challenge Task 2: Unsupervised Anomalous Sound Detection for Machine Condition Monitoring Applying Domain Generalization Techniques. <https://arxiv.org/abs/2206.05876>
- Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., & Lempitsky, V. (2016). Domain-Adversarial Training of Neural Networks. *Journal of Machine Learning Research*, 17(59), 1--35. <https://jmlr.org/papers/v17/15-239.html>
- Gretton, A., Borgwardt, K.M., Rasch, M.J., Schölkopf, B., & Smola, A. (2012). A Kernel Two-Sample Test. *Journal of Machine Learning Research*, 13(25), 723--773. <https://jmlr.org/papers/v13/gretton12a.html>
- Guan, J., Xiao, F., Liu, Y., Zhu, Q., & Wang, W. (2023). Anomalous Sound Detection Using Audio Representation with Machine ID Based Contrastive Learning Pretraining. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. <https://arxiv.org/abs/2304.03588>
- Hinton, G., Vinyals, O., & Dean, J. (2015). Distilling the Knowledge in a Neural Network. *arXiv preprint arXiv:1503.02531*. <https://arxiv.org/abs/1503.02531>
- Howard, A., Sandler, M., Chu, G., Chen, L., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., Le, Q.V., & Adam, H. (2019). Searching for MobileNetV3. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 1314--1324. https://openaccess.thecvf.com/content_ICCV_2019/html/Howard_Searching_for_MobileNetV3_ICCV_2019_paper.html
- Jena, S., Pulkit, A., Singh, K., Banerjee, A., Joshi, S., Ganesh, A., Singh, D., & Bhavsar, A. (2024). Unified Anomaly Detection Methods on Edge Device Using Knowledge Distillation and Quantization. *arXiv preprint arXiv:2407.02968*. <https://arxiv.org/abs/2407.02968> [Vision-domain (image) anomaly detection benchmark; cited here only as a general knowledge-distillation-plus-quantization edge-deployment precedent, not as an audio-domain result.]
- Koizumi, Y., Kawaguchi, Y., Imoto, K., Nakamura, T., Nikaido, Y., Tanabe, R., Purohit, H., Suefusa, K., Endo, T., Yasuda, M., & Harada, N. (2020). Description and Discussion on DCASE2020 Challenge Task2: Unsupervised Anomalous Sound Detection for Machine Condition Monitoring. *arXiv preprint arXiv:2006.05822*. <https://arxiv.org/abs/2006.05822>
- Nishida, T., Harada, N., Niizumi, D., Albertini, D., Sannino, R., Pradolini, S., Augusti, F., Imoto, K., Dohi, K., Purohit, H., Endo, T., & Kawaguchi, Y. (2023). Description and Discussion on DCASE 2023 Challenge Task 2: First-Shot Unsupervised Anomalous Sound Detection for Machine Condition Monitoring. <https://arxiv.org/abs/2305.07828>

- Nishida, T., Harada, N., Niizumi, D., Albertini, D., Sannino, R., Pradolini, S., Augusti, F., Imoto, K., Dohi, K., Purohit, H., Endo, T., & Kawaguchi, Y. (2025). Description and Discussion on DCASE 2025 Challenge Task 2: First-Shot Unsupervised Anomalous Sound Detection for Machine Condition Monitoring. <https://arxiv.org/abs/2506.10097>
- Purohit, H., Tanabe, R., Ichige, K., Endo, T., Nikaido, Y., Suefusa, K., & Kawaguchi, Y. (2019). MIMII Dataset: Sound Dataset for Malfunctioning Industrial Machine Investigation and Inspection. Proceedings of the Detection and Classification of Acoustic Scenes and Events 2019 Workshop (DCASE2019). <https://arxiv.org/abs/1909.09347>
- Saengthong, P., & Shinozaki, T. (2024). Deep Generic Representations for Domain-Generalized Anomalous Sound Detection. <https://arxiv.org/abs/2409.05035>
- Sun, B., & Saenko, K. (2016). Deep CORAL: Correlation Alignment for Deep Domain Adaptation. Computer Vision -- ECCV 2016 Workshops. https://doi.org/10.1007/978-3-319-49409-8_35
- Tan, M., & Le, Q. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. Proceedings of the 36th International Conference on Machine Learning (ICML), 6105--6114. <https://proceedings.mlr.press/v97/tan19a.html>
- Wilkinghoff, K. (2021). Sub-Cluster AdaCos: Learning Representations for Anomalous Sound Detection. Proceedings of the International Joint Conference on Neural Networks (IJCNN). https://dcase.community/documents/challenge2021/technical_reports/DCASE2021_Wilkinghoff_31_t2.pdf
- Wilkinghoff, K. (2024). Self-Supervised Learning for Anomalous Sound Detection. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 276--280. <https://arxiv.org/abs/2312.09578>
- Ye, T., Peng, T., & Yang, L. (2025). Review on Sound-Based Industrial Predictive Maintenance: From Feature Engineering to Deep Learning. Mathematics, 13(11), 1724. <https://doi.org/10.3390/math13111724>
- Yeo, J.J.S., Tan, E., Bai, J., Pekski, S., & Gan, W. (2024). Data Efficient Acoustic Scene Classification Using Teacher-Informed Confusing Class Instruction. <https://arxiv.org/abs/2409.11964> [DCASE 2024 Task 1 (acoustic scene classification, not anomaly detection); cited for its teacher-student knowledge-distillation design for lightweight audio models, which directly parallels LDIAAD's teacher-student distillation (Section 4.3).]