ASD3DP: A Process-Aligned Dataset for Soft and Hard Anomaly Detection in FFF 3D Printing

JeongSik Kim[1,2], JongWoo Sung[1,2], HyeonJun Bae[1,2], BoRyeon Kim[1], JiAn Lee[1]

[1]LUDO Lab, Busan, South Korea [2]Dong-A University, Busan, South Korea

Abstract— Existing anomalous sound detection (ASD) datasets capture industrial noise, domain shifts, and clip-level operating conditions, but provide limited time-aligned evidence for relating detector responses to the evolving operation of a complex machine. We present ASD3DP, a process-aligned operating-sound dataset for fused filament fabrication (FFF) 3D printing. The dataset contains 41 recording sessions and 28.99 h of machine operation acquired through four synchronized microphones, together with 9,067 selected 10 s clips covering normal operation, belt-tension conditions, extruder failures, and toolhead collisions. Each audio interval is linked to a time-aligned command stream, telemetry, process states, print features, and fault-active intervals on a common release-audio timeline. The release supports conventional clip-based ASD and annotation-linked analysis of audio-only model responses. Baseline experiments establish compatibility with reconstruction-based and pretrained-encoder ASD systems. Process attributes are reserved for analysis after the systems and decision thresholds are fixed. The dataset, annotations, and accompanying documentation are publicly available at https://github.com/LUDO-Lab-Research/ASD3DP

Index Terms— Anomalous sound detection, datasets, 3D printing, process annotation, model analysis

1. INTRODUCTION

Anomalous sound detection (ASD) identifies abnormal machine conditions from sound, often using only normal recordings for training. Existing datasets already capture industrial noise, controlled damage, and operating variation: MIMII records industrial machines under factory noise [1]; ToyADMOS and ToyADMOS2 use miniature machines with controlled damage and recording conditions [2], [3]; and MIMII DUE and MIMII DG explicitly address domain shifts [4], [5].

The gap is therefore not simply a lack of 3D-printer audio. DCASE 2024 Task 2 includes this machine type [6]. Related work covers acoustic monitoring [7], print-quality prediction [8], multisensor anomaly data [9], and audio/vibration recordings with design and printer-command files [10]. ASD3DP instead emphasizes linking continuous multichannel audio and anomaly labels to the changing operation of the machine, while retaining a conventional clip-based view.

An FFF printer combines motion, extrusion, cooling, frame vibration, and nozzle-part interaction. Motors, belts, linear-motion components, fans, and the extruder contribute overlapping sounds. A normally operating subsystem can mask another subsystem's fault, and the mixture changes with speed, direction, acceleration, and toolpath. Process annotations help distinguish this normal variation from fault-related responses.

Metadata can also support analysis without becoming a detector input. DCASE distinguishes attribute-available and attribute-unavailable settings [11]; WILDS retains metadata for stratified evaluation [12]; and slice-discovery studies examine coherent error subsets while noting limitations of their interpretation [13]–[15]. Here, session and condition labels define the evaluation subset. Dense process attributes are joined to detector outputs only after model configurations and thresholds are fixed; they are not used to tune or score the detectors.

Our contributions are: (1) continuous four-channel audio and selected 10-s clips from a coupled FFF machine; (2) a common audio timeline linking commands, telemetry, process states, print features, and fault intervals with recorded evidence sources; and (3) analysis showing how these annotations reveal normal-process variation and help interpret fixed audio-only representations and detector responses.

2. DATASET DEFINITION AND SCOPE

ASD3DP contains 41 sessions totaling 28 h 59 min of operation (28.99 h). Four synchronized channels yield 164 session-channel WAV files and 115.97 channel-hours. The clip view contains 9,067 unique 10-s intervals, corresponding to 36,268 channel clips. Table 1 reports the distribution without counting the same operating interval four times.

G-code is the printer command sequence describing motion, extrusion, and control actions; telemetry comprises timestamped printer-state samples. These contextual records complement, rather than replace, audio. The open-source Voron 0 design [16] makes the platform inspectable. Data and annotations are available through the project repository. A detailed collection-protocol repository is planned. This release uses one printer build and does not establish cross-printer generalization.

3. DATA ACQUISITION

3.1. Printer and print protocol

One unenclosed Voron 0.2-class printer was recorded during loaded printing and no-filament motion-only operation. The Voron Design Cube provided a repeatable workload with walls, corners, overhangs, and changing cooling settings. “Slow” and “fast” describe session-level collection protocols, not constant instantaneous speeds. Time-varying motion and extrusion should be analyzed through synchronized telemetry. Release manifests describe the model, material, components, firmware, and protocols.

3.2. Four-channel recording layout

Four Behringer C-2 cardioid condenser microphones were connected to one Zoom AMS-44 and recorded synchronously. CH1, CH2, CH3, and CH4 denote the rear, front, right, and left positions. Each microphone was approximately 5 cm from the nearest printer surface and aimed diagonally toward the build volume (Fig. 1). Separate mono WAV files were stored at 48 kHz and a nominal 768 kb/s per file.

The four positions provide different mixtures of motion, extrusion, cooling, and frame sound; the shared interface preserves relative channel timing without post-hoc resampling. Figure 2 is an oblique interior photograph, not an elevation diagram. It shows the mounting arrangement; Fig. 1 defines the named positions in plan view. No placement-optimization study is claimed. Channel and setup metadata are recorded in audio_channels.csv and session_machine_profile.csv.

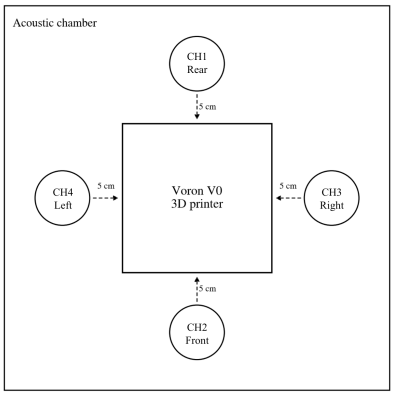

Fig. 1: Plan view of the acquisition setup (not to scale). CH1/rear, CH2/front, CH3/right, and CH4/left are approximately 5 cm from the nearest printer surface. Microphones point toward the build volume.

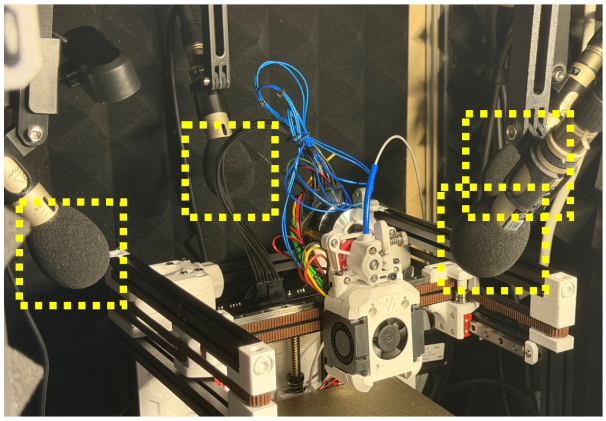

Fig. 2: Oblique interior view of the four microphones (dashed boxes) directed toward the build volume. The plan view in Fig. 1 defines their front, left, right, and rear positions.

Table 1: Session and selected 10-s clip distribution. Durations use h:mm; block headings are subtotals, not additional sessions.

Condition / protocol	Sessions	Session h:mm	Clips	Clip h:mm
Normal subtotal	**19**	**14:01**	**4,766**	**13:14**
Loaded PLA, no fault - slow	12	10:37	3,574	9:56
Loaded PLA, no fault - fast	3	1:26	505	1:24
No filament, no fault - slow	1	0:45	264	0:44
No filament, no fault - fast	3	1:13	423	1:11
Anomaly subtotal	**22**	**14:59**	**4,301**	**11:57**
Belt tension, no filament - slow	3	2:34	900	2:30
Belt tension, no filament - fast	9	4:23	1,452	4:02
Observed blocked path - slow	2	0:55	294	0:49
G-code feed failure - slow	1	0:47	282	0:47
G-code feed failure - fast	3	1:21	476	1:19
G-code toolhead collision, no filament - varied	4	4:58	897	2:30
Total (normal + anomaly)	**41**	**28:59**	**9,067**	**25:11**

Durations are rounded to the nearest minute. Session durations are converted from reported decimal hours; clip durations use counts × 10 s. Independently rounded rows may not sum exactly.

3.3. Conditions and physical definitions

A soft anomaly permits continued operation despite degradation, reduced print quality, or increased failure risk. Current soft conditions are loose A, B, or A+B belts and extruder failures, including an observed blocked filament path and G-code-induced feed failures. A hard anomaly requires immediate intervention or presents collision/damage risk; toolhead collision is the current hard class. "G-code simulation" denotes programmed physical fault induction, not synthesized audio. Observed and induced conditions retain separate provenance. These labels describe operational consequences rather than sound level; an annotated fault condition need not be audible throughout its interval.

3.4. Audio views and contextual records

Full session files preserve continuous operation for custom segmentation. Selected 10-s clips support conventional ASD without imposing a mandatory split. Manifests map each clip and channel to its source interval. Table 2 summarizes the released contextual layers. The v0.4.3 annotation bundle references the audio identities rather than duplicating the WAV files. Command, telemetry, phase, and fault records retain their respective evidence sources.

4. ANNOTATION PROTOCOL

4.1. Audio–printer synchronization

Audio and printer records use different time origins. For each session s, human review established a corresponding print-end anchor in the audio and printer records. Let a_s be its release-audio time and p_s its printer time. The constant-offset mapping is

$$\Delta_s = a_s - p_s, \qquad t_{release} = t_{printer} + \Delta_s. \qquad (1)$$

The session-specific offset and release trims are recorded in session_sync_offsets.csv. One offset is used throughout the session; no clock-drift correction was applied. Command times are estimates obtained by interpolating source G-code byte positions against synchronized telemetry/context anchors. Reconstructed macro-command times preserve logical order but are not exact execution traces. End-to-end alignment error was not independently measured, so temporal correspondence must not be interpreted as sample-accurate machine timing.

Belt and extruder intervals describe synchronized fault-capable ranges; collision intervals come from audio and spectrogram review. The final selection table contains 1,373 fault-active intervals (18 belt/extrusion and 1,355 collision intervals), distinct from diagnostic candidate intervals. One annotator reviewed each collision session; independent double annotation and formal adjudication were not performed. Telemetry and process timelines cover all selected clips, whereas compact G-code context covers 9,055 of 9,067 clips.

4.2. Fixed-grid clip selection

Segmentation and train/test partitioning are separate steps. Within each session, the fixed grid defines non-overlapping intervals [10k, 10(k + 1)) s in release-audio time. The same boundaries identify all four channels. Normal clips are retained only when fully contained in a released printing-motion window, including eligible no-filament motion. Anomaly clips require at least 0.1 s overlap with a final fault-active interval; the release additionally requires full containment for short labeled events. Interval intersections retain process and fault context, and manifests preserve source times. An anomalous clip may therefore include normal-sounding frames.

No window-level random split is imposed on the dataset. The experiment in Section 5 uses disjoint recording sessions, so clips and microphone views from a source session cannot cross between training and testing.

Table 2: Released annotation layers, alignment basis, and coverage. Timing estimates and evidence sources are distinguished from direct acoustic observations.

Annotation layer	Timing / evidence basis	Resolution	Use	Public coverage
Session context	Reviewed acquisition records	Session	Condition, machine, material, protocol, provenance	41 session profiles
Audio alignment	Reviewed anchor offsets and release trims	Session / channel	Common release-audio timeline	41 offsets; 164 WAV files
Command / process	G-code timing estimates and synchronized printer states	Command / interval	Motion, deposition, travel, retraction, feature and layer	3,804,842 commands; 70,912 labels; 78,276 clip-process rows
Telemetry	Timestamped samples mapped to release-audio time	Sample / clip	Motion, extrusion, temperature, fan and progress	2,166,672 samples; 9,067 clip rows
Fault evidence	Protocol-defined ranges; reviewed collision audio	Interval	Fault context and acoustic event analysis	18 belt/extrusion + 1,355 collision selection intervals
Clip view	10-s grid intersected with aligned intervals	Clip / intersection	Conventional ASD and post-hoc analysis	9,067 clips; 36,268 channel rows; 3,693,595 clip-command rows

5. BENCHMARK AND ANALYSIS

The experiments ask three questions: can established audio-only systems use the clip view; what changes occur within valid normal operation; and how do process annotations help interpret fixed model outputs? They demonstrate uses of the resource, not independent proof of annotation accuracy or metadata-assisted detection gains.

5.1. Evaluation subset and audio-only systems

A fixed session-disjoint split (seed 13711) contains 1,000 normal training clips from 16 sessions and 600 test clips: 300 normal clips from three held-out sessions and 100 clips from each fault family. Training normals include 250 clips from each slow/fast loaded-printing and no-filament context. Test normals include 100 clips each from slow loaded printing, fast loaded printing, and fast no-filament motion. Belt and collision tests use no-filament protocols; extruder tests use extrusion-failure protocols. No anomaly is used for training or threshold selection.

We evaluate a DCASE-style autoencoder (AE) [17] and a pretrained-encoder system [18], [19]. The AE uses front-channel audio. “Front-only” and “Rear-only” mean embeddings from CH2 and CH1, respectively; the residual view in Table 4 is the rear embedding minus 0.5 times the front embedding, not waveform subtraction. All systems use audio only. Dense process attributes are inspected after detector configurations, outputs, and decision thresholds are fixed.

We report area under the receiver operating characteristic curve (AUC), partial AUC (pAUC; maximum false-positive rate 0.1), and F1 at train-normal thresholds. AE results are means and standard deviations over three seeds. Its threshold is the 90th percentile of a Gamma fit to final-epoch training-batch losses. The encoder results use their fixed train-normal threshold; Table 4 does not report run-to-run uncertainty.

5.2. What the baseline results show

Table 3: Front-channel AE results, mean (SD) over three seeds. Each fault row compares 300 normal and 100 anomalous clips; the overall row uses all 600 test clips. pAUC: maximum FPR 0.1. F1: Gamma-fit q90 train-normal threshold.

Evaluation scope	AUC	pAUC	F1
Overall	0.796 (0.019)	0.722 (0.009)	0.620 (0.089)
Belt tension	0.723 (0.024)	0.592 (0.016)	0.295 (0.189)
Extruder failure	0.665 (0.051)	0.577 (0.016)	0.290 (0.201)
Toolhead collision	0.998 (0.001)	0.996 (0.003)	0.603 (0.331)

Table 4: Pretrained-encoder results by microphone embedding view on the same test subset. F1 uses the fixed train-normal threshold. The residual is computed in embedding space.

Embedding view	AUC	pAUC	F1
Front-only	0.9657	0.8308	0.8370
Rear-only	0.9873	0.9525	0.9389
Rear − 0.5 front	0.9895	0.9589	0.9416

Table 3 separates overall and fault-family performance. The AE ranks collisions strongly (AUC 0.998), but belt and extruder AUCs are 0.723 and 0.665. Thus, these soft conditions remain harder for this baseline. Collision F1 is 0.603 with SD 0.331: strong ranking does not guarantee stable alarms at a fixed threshold. Table 4 enables a matched-session microphone-view comparison: Rear-only outperforms Front-only, while the residual view improves modestly over Rear-only (AUC 0.9873 to 0.9895). This illustrates the utility of synchronized channels, not a statistically established or universally beneficial fusion rule.

5.3. Retraction as normal acoustic variation

Retraction briefly withdraws filament, unlike deposition, which lays down material. To isolate normal-process variation, we compared 1,133 front-channel 3-s segments containing at least 0.25 s of retraction with nearest-in-time references from the same session and speed profile (no retraction; at least 2.5 s deposition). Standardized differences were averaged within each of six sessions, then equally weighted. Figure 3 shows lower root-mean-square (RMS) level and 0–1 kHz energy ratio, but higher spectral centroid and 1–3/3–8 kHz energy ratios, during retraction. Directions agree across the six sessions. These are differences between valid normal states, illustrating the context lost by a binary normal label; they do not establish that retraction causes false alarms.

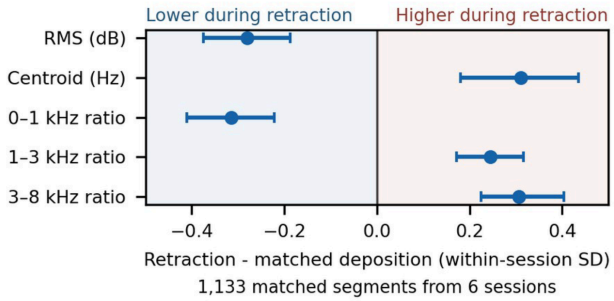

Fig. 3: Normal retraction minus matched deposition, in within-session standard-deviation units. Points equally weight six session means; bars are 95% session-bootstrap intervals. Negative values indicate lower values during retraction.

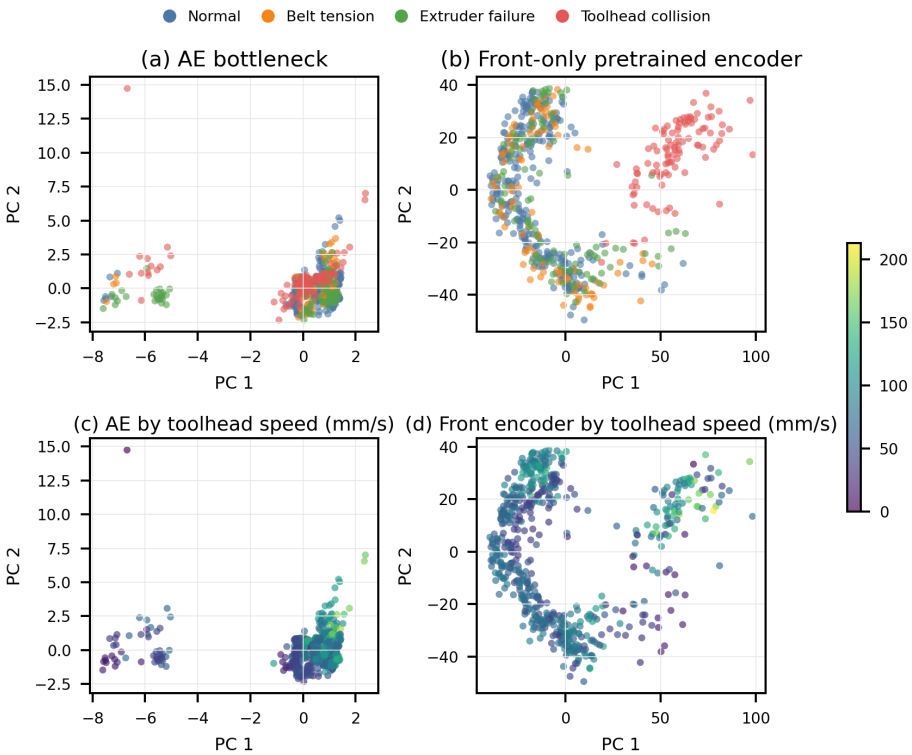

Fig. 4: Separate PCA projections of standardized test representations: AE bottleneck (left), Front-only encoder (right). Top: fault label. Bottom: mean toolhead velocity for the same points. Different projections have different axes.

5.4. Fault and operating-context representations

Figure 4 shows separate principal component analysis (PCA) projections of standardized AE bottleneck and Front-only encoder representations. Top panels show fault labels; bottom panels color the same points by mean toolhead velocity. Collisions are comparatively distinct; normal, belt, and extruder samples overlap. The encoder groups samples differently from the AE, with speed-related structure also visible. Because fault, protocol, and session are not fully crossed, groups may reflect operating conditions as well as faults. These exploratory projections do not establish causal fault features or unbiased sampling; their independently fitted axes are not directly comparable.

5.5. Inspecting temporal detector responses

Figure 5 aligns frame scores with a detected collision and a missed extruder-condition clip. The localized collision response contrasts with fluctuating scores across a broader extruder-condition interval; not every frame in that interval must sound abnormal. The dashed line is the train-normal frame-score 99th percentile, a diagnostic reference, not the file-level decision threshold. These selected examples demonstrate inspectable temporal responses, not measured early-warning accuracy.

6. DISCUSSION AND LIMITATIONS

The release covers one printer, a restricted fault vocabulary, controlled protocols, and coupled condition/session effects. Most test clips are deposition-dominant, fan telemetry is sparse, and retraction analysis covers six sessions. Within-session comparisons do not eliminate fault/speed/protocol confounding. Unmeasured synchronization error and single-annotator collision review further limit temporal ground-truth claims.

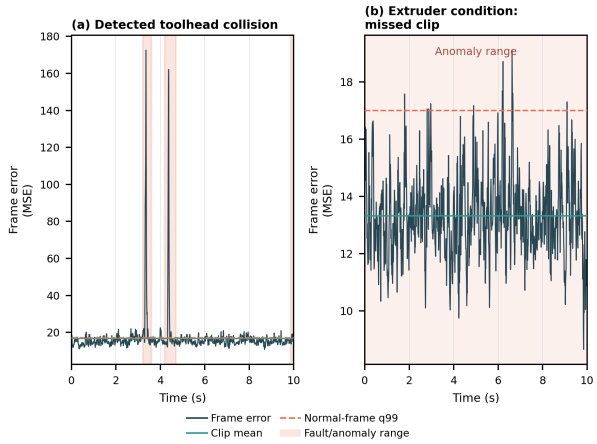

Fig. 5: Frame-score examples aligned with labeled intervals: (a) detected collision; (b) missed extruder condition. Solid line: clip mean. Dashed line: train-normal frame-score 99th percentile, not the file-level decision threshold.

Future releases will add protocol-matched normals and independent Voron V0 builds spanning component manufacturers, printed-part quality, assembly, maintenance, materials, and environments. Planned leave-one-build-out evaluation will tune only on training printers; audio-plus-context studies will include process-only leakage baselines.

7. CONCLUSION

ASD3DP provides four-channel FFF sound and process-aligned annotations for clip-based ASD and interpretation of normal-process, microphone-view, and temporal variation. Current results establish baseline compatibility, not cross-build generalization.

REFERENCES

[1] H. Purohit, R. Tanabe, K. Ichige, T. Endo, Y. Nikaido, K. Suefusa, and Y. Kawaguchi, "MIMII dataset: Sound dataset for malfunctioning industrial machine investigation and inspection," in Proc. DCASE Workshop, 2019, pp. 209-213.

[2] Y. Koizumi, S. Saito, H. Uematsu, N. Harada, and K. Imoto, "ToyADMOS: A dataset of miniature-machine operating sounds for anomalous sound detection," in Proc. IEEE WASPAA, 2019, pp. 313-317.

[3] 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," arXiv preprint arXiv:2106.02369, 2021.

[4] R. Tanabe, H. Purohit, K. Dohi, T. Endo, Y. Nikaido, T. Nakamura, and Y. Kawaguchi, "MIMII DUE: Sound dataset for malfunctioning industrial machine investigation and inspection with domain shifts due to changes in operational and environmental conditions," in Proc. IEEE WASPAA, 2021, pp. 21-25.

[5] 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. DCASE Workshop, 2022.

[6] DCASE Challenge, "DCASE 2024 Challenge Task 2 Additional Training Dataset," Zenodo, 2024.

[7] P. Becker, C. Roth, A. Roennau, and R. Dillmann, "Acoustic anomaly detection in additive manufacturing with long short-term memory neural networks," in Proc. IEEE ICIEA, 2020, pp. 921-926.

[8] M. Olowe, M. Ogunsanya, B. Best, Y. Hanif, S. Bajaj, V. Vakkalagadda, O. Fatoki, and S. Desai, "Spectral features analysis for print quality prediction in additive manufacturing: An acoustics-based approach," Sensors, vol. 24, no. 15, p. 4864, 2024.

[9] J. Sendorek, T. Szydlo, M. Windak, and R. Brzoza-Woch, "Dataset for anomalies detection in 3D printing," in Computational Science - ICCS 2021, ser. Lecture Notes in Computer Science, vol. 12746, 2021, pp. 647-653.

[10] C. Madamopoulos and N. G. Tsoutsos, "3D printer audio and vibration side channel dataset for vulnerability research in additive manufacturing security," Data in Brief, vol. 57, p. 111002, 2024.

[11] T. Nishida et al., "Description and discussion on DCASE 2024 Challenge Task 2: First-shot unsupervised anomalous sound detection for machine condition monitoring," arXiv preprint arXiv:2406.07250, 2024.

[12] P. W. Koh, S. Sagawa, H. Marklund, S. M. Xie, M. Zhang, A. Balsubramani, W. Hu, M. Yasunaga, R. L. Phillips, I. Gao, T. Lee, E. David, I. Stavness, W. Guo, B. Earnshaw, I. Haque, S. M. Beery, J. Leskovec, A. Kundaje, E. Pierson, S. Levine, C. Finn, and P. Liang, "WILDS: A benchmark of in-the-wild distribution shifts," in Proc. ICML, ser. Proceedings of Machine Learning Research, vol. 139, 2021, pp. 5637-5664.

[13] S. Eyuboglu, M. Varma, K. Saab, J.-B. Delbrouck, C. Lee-Messer, J. Dunnmon, J. Zou, and C. Ré, "Domino: Discovering systematic errors with cross-modal embeddings," in Proc. ICLR, 2022.

[14] Y. Yang, H. Zhang, D. Katabi, and M. Ghassemi, "Change is hard: A closer look at subpopulation shift," in Proc. ICML, 2023, arXiv:2302.12254.

[15] N. Johnson, Á. A. Cabrera, G. Plumb, and A. Talwalkar, "Where does my model underperform? a human evaluation of slice discovery algorithms," in Proc. AAAI Conf. Human Computation and Crowdsourcing, vol. 11, no. 1, 2023, pp. 65-76.

[16] VoronDesign, "Voron 0 CoreXY 3D printer design," GitHub repository, accessed: Jul. 2026. [Online]. Available: https://github.com/VoronDesign/Voron-0.

[17] N. Harada, D. Niizumi, Y. Ohishi, D. Takeuchi, and M. Yasuda, "First-shot anomaly sound detection for machine condition monitoring: A domain generalization baseline," in Proc. EUSIPCO, 2023, pp. 191-195.

[18] T. Alex, S. Ahmed, A. Mustafa, M. Awais, and P. J. B. Jackson, "SSLAM: Enhancing self-supervised models with audio mixtures for polyphonic soundscapes," in Proc. ICLR, 2025.

[19] J. Kim, J. Sung, H. Bae, B. Kim, and J. Lee, "Residual view and prototype selection for noise-aware anomalous sound detection," DCASE 2026 Challenge, Tech. Rep., Jun. 2026.