IMAD-DS: A DATASET FOR INDUSTRIAL MULTI-SENSOR ANOMALY DETECTION UNDER DOMAIN SHIFT CONDITIONS

Davide Albertini[1], *Filippo Augusti* [1], *Kudret Esmer*[2], *Alberto Bernardini*[2], *Roberto Sannino*[1]

[1] STMicroelectronics, Agrate Brianza (BG), Italy, {davide.albertini, filippo.augusti,roberto.sannino}@st.com
[2] Politecnico di Milano, DEIB, Milano (MI), Italy,
kudret.esmer@mail.polimi.it, alberto.bernardini@polimi.it

ABSTRACT

Industrial anomaly detection (AD) plays a critical role in maintaining the safety, efficiency and productivity of modern manufacturing and production processes. Despite the widespread adoption of IoT sensor boards in industry, there is still a lack of comprehensive multi-sensor and multi-rate datasets for AD that adequately account for domain shifts, i.e. variations in operational and environmental conditions that significantly affect AD performance. To address this gap, we present the Industrial Multi-sensor Anomaly Detection under Domain Shift Conditions (IMAD-DS) dataset. The IMAD-DS dataset comprises multi-sensor data from two scaled industrial machines: a robotic arm and a brushless motor, collected under different operating conditions to mimic real-world domain shifts, including speed and load changes. We also add different types of background noise to the audio data to simulate different environmental domain shifts. Benchmark testing with an autoencoder model show that AD performance decreases significantly with domain shifts, emphasizing the value of IMAD-DS for the development of robust multi-sensor AD systems.

Index Terms— Anomaly Detection, Sensor Fusion, Dataset, Domain Shift

1. INTRODUCTION

As modern industry grows in complexity and scale, the role of anomaly detection (AD) in machine monitoring and fault detection has increased significantly. This brings several benefits, such as increased safety, reduced impact on machine performance and higher productivity. Traditionally, industrial AD has relied on the experience of on-site technicians. While effective, this method is labor-intensive and often limited by the physical accessibility of some machine components. Therefore, the shift towards automated, data-driven methods such as machine learning and deep learning has gained momentum [1]. In this context, AD is framed as the task of automatically detecting abnormal conditions by learning only normal operating conditions.

A variety of physical variables such as vibration [2, 3, 4], temperature [5], pressure [6], and audio [7, 8, 9] can be used to detect anomalies in the industrial environment. However, with the widespread adoption of IoT boards it is now possible to simultaneously collect data from numerous sensors, providing a more comprehensive multi-modal description of machine operation. This data enables the development of more robust AD algorithms that take advantage of this richer description. Thus, the presence of multi-modal AD datasets becomes crucial for the development of the next generation of data-driven industrial AD systems.

Nevertheless, most existing industrial AD datasets primarily focus on single-sensor data, with only a few datasets covering multi-sensor scenarios. Notably, the Tennessee Eastman Process (TEP) models an industrial chemical process using a model-based simulator [10]. The HAI dataset captures data from a realistic industrial control system augmented with a hardware-in-the-loop simulator [11]. The CWRU Bearing dataset focuses on motor condition assessment [12]. Additionally, the Skoltech Anomaly Benchmark (SKAB) provides data from various machines captured using multiple sensors [13]. However, these datasets often overlook the inherent variability of real industrial environments that significantly affect the performance of AD systems [14, 15, 16, 17]. These deviations are often referred to as domain shifts and represent natural deviations in the distribution of normal data, which, however, make the automatic detection of anomalies more difficult.

The importance of accounting for domain shifts has recently been recognized in the field of audio-based anomaly detection, thanks in part to the contributions of the DCASE Task2 challenge and the availability of datasets that take this aspect into account, such as TOYADMOS2 [15], MIMII DUE [16] and MIMII-DG [17]. Introducing domain shifts into a dataset enables the development of more robust AD models and facilitates the development of domain adaptation and generalization techniques [17].

Inspired by the growing interest for AD in the presence of domain shifts, this paper introduces the Industrial Multi-sensor Anomaly Detection under Domain Shift Conditions (IMAD-DS) dataset. IMAD-DS comprises multi-sensor data from two scaled representations of industrial machines, namely a robotic arm and a brushless motor, collected under varying operational conditions to mimic real-world domain shifts, which include variations in operating speeds and loads. We also add different types of background noise to the audio signals to simulate different environmental domain shifts. Further, the dataset comprises sensors producing data with different sampling frequency, increasing the complexity with respect to single-rate multi-sensor datasets such as [10, 11, 13].

In addition to the dataset, we propose a deep learning model that enables multi-modal and multi-rate anomaly detection (AD) under domain shift conditions, serving as a benchmark to evaluate the dataset's usefulness. The model employs a fully connected autoencoder (AE) architecture that attempts to reconstruct multi-sensor data, yielding a reconstruction error which serves as an anomaly score metric for unsupervised AD. Results show that using multiple sensors is helpful for the task of AD, and also that performance decreases under domain shifts, underscoring the usefulness of the IMAD-DS dataset. The dataset is freely available for download at `https://zenodo.org/records/12636236`.

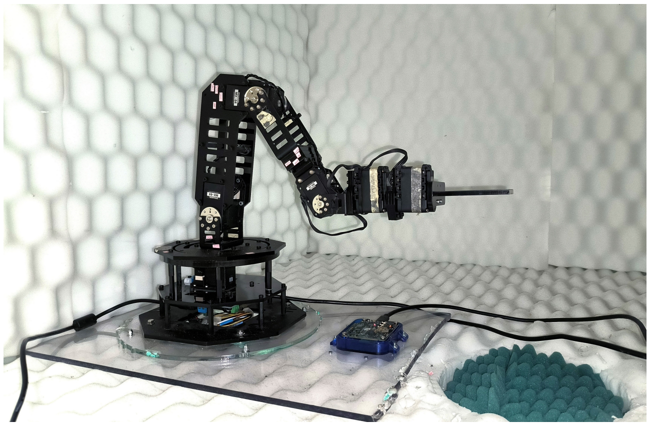

Figure 1: Robotic arm in the anechoic chamber, without weights. The IoT acquisition board is connected to the machine through the plexiglass base.

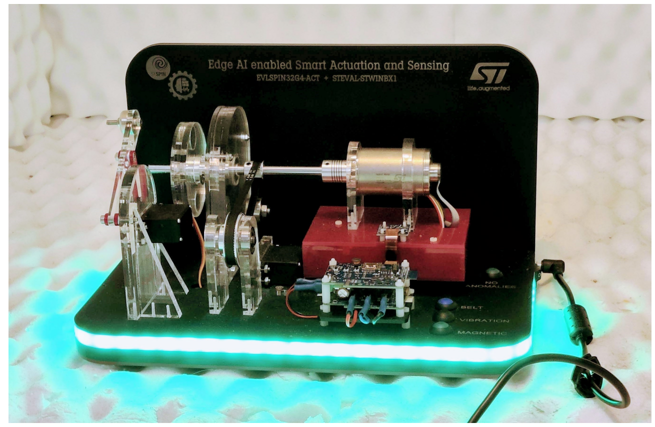

Figure 2: Brushless motor in the anechoic chamber. The IoT acquisition board is connected to the machine by screws that hold it on the plastic base.

Machine Name	Domain Shift Parameter	Value for Source-Domain	Value for Target-Domain
Robotic Arm	Attached loads of increasing weight	00, 10, 15, 20	25, 30, 35
	Factory background noise (SNR -4 dB)	A, C, D, F	B, E, G
Brushless Motor	Different rotation speeds [rpms]	1500, 1600, 1700, 1800, 1900, 2000, 2400, 2800, 3000	1000, 1100, 1200, 1300, 1400
	Factory background noise (SNR -4 dB)	A, C, D, F	B, E, G

Table 1: Domain shift configurations for the robotic arm and brushless motor in the IMAD-DS dataset. The table lists the different operational and environmental conditions used to create source and target domains. Numbers from 00 to 35 are indexes of increasing weights. Letters A to F index 7 background noise recordings from different real factories, all scaled to attain a SNR of -4 dB.

	Robotic Arm				Brushless Motor			
	Source Domain		Target Domain		Source Domain		Target Domain	
	Normal	Anomaly	Normal	Anomaly	Normal	Anomaly	Normal	Anomaly
Train	1812	0	27	0	1263	0	18	0
Test	116	116	116	116	78	78	78	78

Table 2: Number of samples for each class in source and target domains, further divided into normal and anomaly classes for the two machines.

2. DATASET OVERVIEW

The IMAD-DS dataset comprises multi-rate and multi-sensor data from two scaled representations of industrial machines, namely a robotic arm and a brushless motor. It contains both normal and abnormal multi-sensor data, which are also recorded under different operating conditions to account for domain shifts. The domain shifts considered in this dataset are divided into operational domain shifts, which are all the allowed machine working configurations, and environmental domain shifts, which are caused by changes in background noise. Anomalies are introduced by intentional disruptions to the normal behavior of the machine in question. IMAD-DS dataset consider the following machines.

Robotic Arm: The robotic arm is a scaled version of a robotic arm used to move silicon wafers in a factory, reproducing actual factory movements. The machine and its recording setup are shown in Fig. 1. Anomalies are created by loosening the screws at the arm's nodes, causing the typical spatial miscalibrations of such machines.

Brushless Motor: The brushless motor is a scaled representation of an industrial brushless motor, as shown in Fig. 2. Two anomalies are introduced: first, a magnet is moved closer to the motor load, causing oscillations by interacting with two symmetrical magnets on the load; second, a belt that rotates in unison with the motor shaft is tightened, creating mechanical stress.

To introduce domain shifts, various operating and environmental conditions are considered for each machine type. The robotic arm is recorded with seven different loads of increasing weight. In contrast, the brushless motor is recorded using 14 different operating voltages leading to various speeds. Both machines are also subjected to different background noises as environmental conditions. Combinations of these operating and environmental conditions divide each machine's dataset into two subsets, namely the source domain and the target domain. The source domain represents the original environment where a large number of training examples are available. In contrast, the target domain is characterized by a series of domain shifts where the availability of training data is severely limited and often restricted to few clips of target condition. The discrepancy between the source and target domains reflects a common problem in practice, where sufficient training data is often not available for the target domain. The domain shift configurations for both datasets are shown in Tab. 1.

As the dataset is tailored for unsupervised anomaly detection,

Sensor Type	Sample Rate	Part Number
Analog microphone	16 kHz	IMP23ABSU
3-axis Accelerometer	6.7 KHz	ISM330DHCX
3-axis Gyroscope	6.7 KHz	ISM330DHCX

Table 3: Sensors embedded on the STWIN.box IoT board and used for data acquisition.

this characteristic is also mirrored in the dataset's composition. Unsupervised AD systems exclusively use normal data for training, since acquiring a comprehensive set of real-world anomalies is challenging. Anomalous samples are included only in the test set to assess the system's capability to detect unknown anomalies. The composition of each machine dataset is detailed in Tab. 2.

3. RECORDING SETUP AND DATA PROCESSING

Multi-sensor data is collected using a STEVAL-STWINBX1 [18], an IoT Sensor Industrial Node from STMicroelectronics. In both machines, the sensor board and the machine lie on the same surface, allowing us to jointly characterize the machine's behavior in terms of audio, vibration, and rotations. The MEMS sensors used to capture these physical quantities are a microphone, an accelerometer, and a gyroscope, respectively. The actual sensors embedded on the sensor board and used to collect data are listed in Tab. 3 along with their respective sampling frequencies.

All recordings are conducted in a completely anechoic chamber, allowing precise control of the acoustic environment. This configuration not only enables detailed acoustic simulations, but also provides the flexibility to adjust the level of background noise to achieve the desired signal-to-noise ratio (SNR) and thus adjust the difficulty of the audio part of the AD task.

3.1. Processing of Audio Signals

The audio signals collected by the microphone are processed to simulate environmental domain shifts. For this purpose, machine noises are mixed with background noises recorded in real factories according to specific SNRs. In order to make the machine sounds and the background noise acoustically coherent, an acoustic simulation is performed to simulate a virtual acoustic environment in which sound sources, i.e. the background noises and the machine sounds, and a virtual microphone are present. Fig. 3 shows the configuration of the virtual acoustic environment in which the background noise sources are placed at the corners of a shoebox room with dimensions $10 \times 7.5 \times 4$ meters. The acoustic simulation is performed by employing the image source method (ISM) [19], which is used to calculate the room impulse response (RIR) of each virtual source and the virtual microphone, thus modeling the multi-path propagation of sound sources in the reverberant environment. In particular, the Pyroomacoustics library [20] is used to implement the ISM and to obtain the RIRs with a fixed reverberation time of $T60 = 0.5$ s.

Given the static nature of the acoustic environment under consideration, the RIRs are computed once for the entire dataset. The subsequent audio processing steps are as follows:

- Selection and Cropping: A background noise signal $\mathbf{n}$ is selected and cropped to match the length of the anechoic machine sound.
- Reverberation of Background Noise: The background noise signal $\mathbf{n}$ is convolved with the RIRs of the background noise emitters, producing the reverberated background noise signal at the virtual microphone $\mathbf{n}^{\text{rev}}$.
- Reverberation of Machine Sound: The machine sound $\mathbf{x}_{\text{mic}}^{\text{anech}}$ is convolved with its corresponding RIR, yielding $\mathbf{x}_{\text{mic}}^{\text{rev}}$.
- Scaling for SNR: The background noise $\mathbf{n}^{\text{rev}}$ is scaled to achieve the desired SNR using

$$\mathbf{n}^{\text{scaled}} = \mathbf{n}^{\text{rev}} \sqrt{\frac{P_{\mathbf{x}_{\text{mic}}^{\text{rev}}}}{10^{\text{SNR}/10} P_{\mathbf{n}^{\text{rev}}}}} \quad (1)$$

  where $P_{\mathbf{n}^{\text{rev}}}$ and $P_{\mathbf{x}_{\text{mic}}^{\text{rev}}}$ denote the power of the reverberated background noise and machine sound, respectively. SNRs are set according to Tab. 1.
- Final Cropping and Mixing: The signals $\mathbf{x}_{\text{mic}}^{\text{rev}}$ and $\mathbf{n}^{\text{scaled}}$ are cropped to the original machine sample length to remove the reverberation tail and are then mixed to produce the final audio sample used in the dataset.

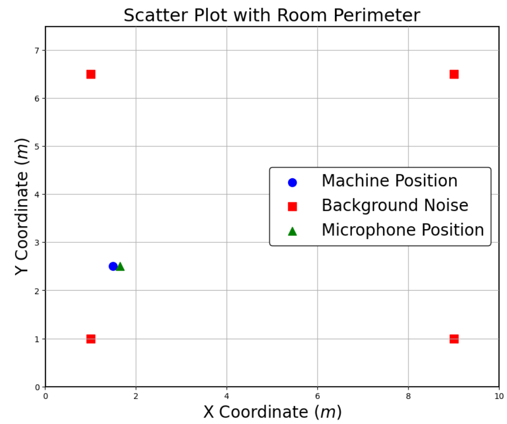

Figure 3: Acoustic environment simulated with the ISM. The red squares indicate the position of the background noise emitters, the blue circle the position of the machine sound emitter and the green triangle the position of the virtual microphone that senses the multi-path propagation of the sound sources.

Note that, in this work, we assume that the coupling of the machine with its surrounding environment is reflected only in the audio signals, as the acoustic coupling is more relevant than the others. The same setup used for the IMADS-DS dataset has also been used for generating audio files for the DCASE2024 task2 challenge *First-Shot Unsupervised Anomalous Sound Detection for Machine Condition Monitoring*.

4. EVALUATION AND BENCHMARK

To give an idea of the use and usefulness of the IMAD-DS dataset, we tested each machine sub-dataset on a simple baseline system. The Python codes for training, testing and creating the training and test data are available in the IMAD-DS dataset public repository.

4.1. Baseline

As a benchmark system, we use a fully-connected autoencoder (AE) that attempts to reconstruct an input vector consisting of all multi-rate, multi-sensor data related to the same temporal window. When an anomalous input is presented, a larger reconstruction error is expected, making the reconstruction error a valid anomaly score metric for unsupervised AD. The input of the baseline system consists of a column vector obtained by concatenating 100 ms windows of multi-sensor data. We denote each sensor data as $\mathbf{x}_s \in \mathbb{R}^{L_s C_s}$, where $s \in \mathcal{S} \triangleq \{\text{mic}, \text{acc}, \text{gyr}\}$ denotes a specific sensor, L_s is the number of samples in the 100 ms window given the sensor's sampling frequency, and C_s is the number of channels for that sensor (e.g., the accelerometer has x-, y- and z- axis components). Note that we stack all the sensor channels to form a single column vector of size $\sum_{s\in\mathcal{S}} L_s C_s$. Moreover, we apply a z-score normalization for each sensor channel, thereby obtaining the normalized sensor data $\tilde{\mathbf{x}}_s \in \mathbb{R}^{L_s C_s}$. Finally, the input of the AE is expressed as the concatenation of each normalized and stacked sensor data, i.e., $\mathbf{x} = [\tilde{\mathbf{x}}_{\text{mic}}^T, \tilde{\mathbf{x}}_{\text{acc}}^T, \tilde{\mathbf{x}}_{\text{gyr}}^T]^T$. The model encoder $E(\cdot|\theta_e)$, defined by trainable parameters θ_e, is composed of 3 fully connected (FC) layers with ReLU activation function [21], namely $FC(\sum_{s\in\mathcal{S}} L_s C_s, 2048, \text{ReLU})$, $FC(2048, 2048, \text{ReLU})$ and $FC(2048, 2048, \text{ReLU})$, and a bottleneck layer $FC(2048, 16, \cdot)$. The decoder $D(\cdot|\theta_d)$ mirrors the architecture of the encoder, with parameters θ_d. The model output is therefore the reconstructed input vector $\mathbf{x}' = D(E(\mathbf{x}|\theta_e)|\theta_d)$. The parameters of the encoder and decoder neural networks (i.e., $\theta = (\theta_e, \theta_d)$) are trained to minimize the loss function given as

$$\mathcal{L}(\theta_e, \theta_d) = \frac{1}{\sum_{s\in\mathcal{S}} L_s C_s} \|\mathbf{x} - D(E(\mathbf{x}|\theta_e)|\theta_d)\|_2^2 \qquad (2)$$

We trained the model with the Adam optimizer [22] with a learning rate of 10^{-4} and a batch size of 1024.

4.2. Results

To assess the AD performance of our benchmark model, we use the Area Under the Receiver Operating Characteristic Curve [23], i.e.,

$$\text{AUC} = \frac{1}{N_d^- N_d^+} \sum_{i=1}^{N_d^-} \sum_{j=1}^{N_d^+} \mathcal{H}\left(A_\theta(\mathcal{X}_j^+) - A_\theta(\mathcal{X}_i^-)\right), \qquad (3)$$

where $\mathcal{X}_i^-$ represents the ith normal data segment from the set of normal test data segments $\{\mathcal{X}_i^-\}_{i=1}^{N_d^-}$, and $\mathcal{X}_j^+$ is the jth anomalous data segment from the set of anomalous test data segments $\{\mathcal{X}_j^+\}_{j=1}^{N_d^+}$. Each data segment consists of several 100 ms windows. In this context, N_d^- and N_d^+ denote the total number of normal and anomalous test segments, respectively, with $d \in \{\text{Source}, \text{Target}, \text{Source} + \text{Target}\}$ specifying the domain under consideration. The anomaly score $A_\theta(\cdot)$ of each segment is the median reconstruction error of all inputs $\mathbf{x}$ within the segment. The function $\mathcal{H}(\cdot)$ outputs 1 if its input is positive, and 0 otherwise. Tab 4 summarizes the AD performance of the benchmark system. The columns labeled 'Source', 'Target', and 'S + T' present the AUC metrics for the source domain, the target domain, and the combined domain, respectively. The 'Overall' row displays the AUC calculated using the anomaly score from (2). To assess the benefits of using multi-sensor data over a single sensor setup, we also set all but one sensor data to zero. For instance, to evaluate the AD performance with only the microphone, we use as input to the AE the vector $\mathbf{x} = [\tilde{\mathbf{x}}_{\text{mic}}^T, \mathbf{0}^T, \mathbf{0}^T]^T$. For this configuration, (2) is evaluated on the subvectors corresponding to the microphone data only, i.e., $\mathbf{x}[: L_{\text{mic}} C_{\text{mic}}]$ for the input and $\mathbf{x}'[: L_{\text{mic}} C_{\text{mic}}]$ for the reconstructed output. The corresponding AUC metric is denoted in Tab. 4 as 'S-mic'. Moreover, we can also evaluate the single sensor AD performance when the others sensor data is present, i.e., when using as input to the AE $\mathbf{x} = [\tilde{\mathbf{x}}_{\text{mic}}^T, \tilde{\mathbf{x}}_{\text{acc}}^T, \tilde{\mathbf{x}}_{\text{gyr}}^T]^T$. For instance, to evaluate how AD performance of just the microphone is influenced by other sensors, we use again (2) on the microphone subvectors. The corresponding AUC is denoted in Tab. 4 as 'F-mic'. Results indicate that

Machine	Robotic Arm			Brushless Motor		
Domain	S + T	Source	Target	S + T	Source	Target
Overall	**91.62**	93.28	90.48	58.95	73.63	55.59
F-acc	90.49	**98.98**	**94.00**	**69.30**	**77.80**	**59.62**
S-acc	88.96	98.40	84.24	67.38	77.17	56.03
F-gyr	87.88	93.91	93.37	57.27	68.28	55.70
S-gyr	46.79	44.99	48.54	57.38	68.11	56.49
F-mic	66.31	73.27	63.18	54.19	58.83	49.27
S-mic	50.92	52.11	49.69	50.71	53.13	46.10

Table 4: Baseline AUC results, in percentage.

sensor-specific AUCs generally improve when incorporating data from other sensors, rather than relying solely on their own data. This suggests that multi-sensor data enhances performance even for single-sensor AD tasks. Furthermore, superior performance in the source domain over the target domain suggests domain shifts pose a challenge in the IMAD-DS dataset. In some instances, the 'S + T' AUC is lower than that of individual domains, as seen with 'F-acc' AUCs for the Robotic Arm dataset. This occurs when normal samples in the target domain have higher anomaly scores than anomalous samples in the source domain, leading the model to mistake domain changes for anomalies. For the Robotic Arm, the 'Overall' AUC exceeds individual sensors' AUCs in 'S + T' domain, which is not the case for the Brushless Motor dataset. This suggests that while sensor data fusion often aids AD, using (2) as an anomaly score does not guarantee that using all sensors always yield optimal performance. Therefore, exploring alternative multi-sensor methods is key to fully exploiting the potential of multi-sensor data.

5. CONCLUSIONS

We presented IMAD-DS, a dataset developed to support the creation of domain adaptation and generalization strategies specifically tailored for multi-rate, multi-sensor AD systems in industrial settings. IMAD-DS includes both normal and abnormal operational data from two scaled versions of industrial machines, each collected under different operational scenarios to account for the variability in the domain. Our experiments with a fusing AE show improvements in AD when data from multiple sensors are included, compared to using data from a single sensor. Furthermore, we observe a decrease in AD efficacy due to domain shifts. This emphasizes the crucial role of IMAD-DS in the development of robust multi-rate multi-sensor systems for AD.

6. ACKNOWLEDGEMENTS

The authors whish to thank Alessandro Ilic Mezza for his support and suggestions throughout the development of the work.

7. REFERENCES

[1] L. Ruff, J. Kauffmann, R. Vandermeulen, G. Montavon, W. Samek, M. Kloft, T. Dietterich, and K. R. Müller, "A unifying review of deep and shallow anomaly detection," *Proceedings of the IEEE*, vol. PP, pp. 1–40, 02 2021.

[2] E. Carden, "Vibration based condition monitoring: A review," *Structural Health Monitoring*, vol. 3, pp. 355–377, 12 2004.

[3] M. Yu, D. Wang, and M. Luo, "Model-based prognosis for hybrid systems with mode-dependent degradation behaviors," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 1, pp. 546–554, 2014.

[4] G. S. Galloway, V. M. Catterson, T. Fay, A. Robb, and C. Love, "Diagnosis of tidal turbine vibration data through deep neural networks," vol. 3, no. 1, Jul. 2016.

[5] G. Lodewijks, W. Li, Y. Pang, and X. Jiang, "An application of the iot in belt conveyor systems," vol. 9864, 09 2016, pp. 340–351.

[6] R. F. Salikhov, Y. P. Makushev, G. N. Musagitova, L. U. Volkova, and R. S. Suleymanov, "Diagnosis of fuel equipment of diesel engines in oil-and-gas machinery and facilities," *AIP Conference Proceedings*, vol. 2141, no. 1, p. 050009, 08 2019. [Online]. Available: https://doi.org/10.1063/1.5122152

[7] H. Hojjati and N. Armanfard, "Self-supervised acoustic anomaly detection via contrastive learning," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, pp. 3253–3257.

[8] Y. Liu, J. Guan, Q. Zhu, and W. Wang, "Anomalous sound detection using spectral-temporal information fusion," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, pp. 816–820.

[9] K. Wilkinghoff, "Utilizing sub-cluster adacos for anomalous sound detection under domain shifted conditions," 2021.

[10] G. Manca, ""tennessee-eastman-process" alarm management dataset," 2020. [Online]. Available: https://dx.doi.org/10.21227/326k-qr90

[11] H.-K. Shin, W. Lee, J.-H. Yun, and H. Kim, *HAI 1.0: HIL-Based Augmented ICS Security Dataset*. USA: USENIX Association, 2020.

[12] W. A. Smith and R. B. Randall, "Rolling element bearing diagnostics using the case western reserve university data: A benchmark study," *Mechanical Systems and Signal Processing*, vol. 64-65, pp. 100–131, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0888327015002034

[13] I. D. Katser and V. O. Kozitsin, "Skoltech anomaly benchmark (skab)," https://www.kaggle.com/dsv/1693952, 2020.

[14] Y. Kawaguchi, K. Imoto, Y. Koizumi, N. Harada, D. Niizumi, K. Dohi, R. Tanabe, H. Purohit, and T. Endo, "Description and discussion on dcase 2021 challenge task 2: Unsupervised anomalous sound detection for machine condition monitoring under domain shifted conditions," 2021. [Online]. Available: https://arxiv.org/abs/2106.04492

[15] 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," 2021.

[16] 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," 2021.

[17] 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," 2022.

[18] "Stwin.box - sensortile wireless industrial node development kit." [Online]. Available: https://www.st.com/en/evaluation-tools/steval-stwinbx1.html

[19] J. Allen and D. Berkley, "Image method for efficiently simulating small-room acoustics," *The Journal of the Acoustical Society of America*, vol. 65, pp. 943–950, 04 1979.

[20] R. Scheibler, E. Bezzam, and I. Dokmanić, "Pyroomacoustics: A python package for audio room simulation and array processing algorithms," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 351–355.

[21] K. Jarrett, K. Kavukcuoglu, M. Ranzato, and Y. LeCun, "What is the best multi-stage architecture for object recognition?" in *2009 IEEE 12th International Conference on Computer Vision*, 2009, pp. 2146–2153.

[22] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization." in *3rd International Conference for Learning Representations (ICLR), San Diego*, Y. Bengio and Y. LeCun, Eds., 2015. [Online]. Available: http://dblp.uni-trier.de/db/conf/iclr/iclr2015.html#KingmaB14

[23] 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," 05 2022.