CANDAS Dataset: A Cooling Fan Sound Dataset with Modeled Disturbances and Controlled Experimental Conditions

Henrique Lima[1], Cristofer Silva[2], Ricardo Nery[1], Wilmer Córdoba Camacho[1], Ricardo Brazileiro[5], Rodrigo de Paula Monteiro[2], Andrea Maria Nogueira Cavalcanti Ribeiro[3], Carmelo José Albanez Bastos-Filho[2], Mariela Cerrada[4], Diego Pinheiro[1]

[1] Universidade Católica de Pernambuco (UNICAP)

[2]Universidade de Pernambuco (UPE)

[3]Universidade Federal de Pernambuco (UFPE)

[4]Universidad Estatal de Milagro

[5]Instituto SENAI

{henrique.00000848801, ricardo.00000851500, wilmer.camacho, diego.silva}@unicap.br, cgss@ecomp.poli.br, ricardo.brazileiro@sistemafiepe.org.br, {rodrigo.monteiro, carmelofilho}@poli.br, andrea.marianogueira@ufpe.br, mcerradal@unemi.edu.ec

Abstract. *Predictive maintenance cuts economic and safety risks in rotating machinery by leveraging vibration and acoustic data, which machine-learning models convert into intelligent fault detectors. Acoustic signals are especially powerful for early fault detection and cooling fans, with simple rotational dynamics, are convenient proxies for complex rotors. Yet existing fan datasets lack disturbance models and controlled conditions. We present CANDAS, a controlled sound dataset featuring 28 h of recordings from two cooling fans under five modeled disturbance conditions. Baseline experiments with three anomaly-detection models validate its value, advancing reproducible research on acoustic fault detection in rotating machinery.*

1. Introduction

Machine failure in the industrial sector can lead to social, economic, and environmental issues, jeopardizing worker safety, economic stability, and product availability for society [Chinniah 2015]. As a result, machinery maintenance has prioritized predictive approaches to avoid unexpected failures [Moura Filho et al. 2023]. Anomaly detection, with the advancements in machine learning and the industry's growing demands, has been widely explored in predictive maintenance by identifying abnormal machine behavior [Chandola et al. 2009]. Machine learning models use sensor data, such as vibration, temperature, pressure, and acoustic signals—acoustic emissions and sound—to identify deviations from expected behavior and classify them as anomalous data. Acoustic signals, when compared with others sensors data, have enabled early anomaly detection. [Saufi et al. 2019].

Although the use of acoustic data in machine learning approaches is very important for the anomaly detection task in rotating machines, there is a notable lack of acoustic datasets that provide diverse and detailed information on observed rotating machines and data acquisition processes. Including detailed information—such as voltage levels, operational loads, and specific configurations related to anomalous behavior—improves the utility of the data and supports fine-grained diagnostics. Clearly understanding the details of the experimental setup and anomaly conditions is essential to improve the explainability of the data set and ensure transparency in the data used with machine learning approaches, fostering the advancement of solutions in the anomaly detection task.

Ensuring components are at the correct temperature is of great value to any system in different contexts and applications. For this purpose, cooling fans are widely used in computational systems and other contexts. They are designed by combining metal heat sinks with a propeller to prevent high temperatures, dissipating the heat generated by components such as processors and power supplies [Al-Hazmi 2020a]. In addition to computing, rotating machinery, such as cooling fans, plays a key role in a variety of applications, including the automotive, industrial, and refrigeration sectors. Due to these widespread applications, advances in the research of anomaly detection in rotating machines are very important to enhance their usefulness and useful life.

In this work, we present the creation of the CANDAS dataset, which contains sounds captured from cooling fans, a simpler instance of rotating machinery that shares fundamental characteristics with more complex systems such as drone propellers, industrial fans, and wind turbines. Additionally, we propose a vibratory disturbance model to support controlled experimentation and analysis. To corroborate our data, we used the Multiple Time Series Analysis Framework (MTSA) [Silva et al. 2024]. The framework has different state-of-the-art machine learning approaches, and we applied them to our proposed dataset, performing structured validation in in-distribution (ID) and out-of-distribution (OoD) scenarios. These two scenarios allow us to validate the robustness of the models and how our data set works with them. The contribution of this paper is threefold. First, we created an open dataset containing sound data from commonly used types of cooling fans, employing an easy-to-reproduce and low-cost setup. Second, we employed a vibratory disturbance model that introduces a range of disturbances with varying levels of severity. Finally, we investigated different machine learning approaches using our dataset to evaluate the performance of these models in detecting anomalies in cooling fans.

2. Related Works

2.1. Anomaly detection in acoustic data

The anomaly detection has been widely explored in different contexts, such as fraud detection [Roy and George 2017], Mars Science Laboratory (MSL) [Zhou et al. 2022], and wind turbines [Roelofs et al. 2024]. In this task, most of the data is used to define the characteristics of normal behavior, and what does not follow these characteristics is labeled as abnormal. The multiple time series is commonly used in this task due to the number of domains in which they appear, such as sensor data [Li et al. 2018] and health data. Given $\mathcal{X}$, where $\mathcal{X} = \{X^1, X^2, ..., X^N\}$, the multiple time series is $\mathcal{X}$ that contains a set of N time series, in this case $X^1, X^2, ..., X^N$ are the constituent series of the $\mathcal{X}$. Whether due to the number of acoustic sensors or data resources, several acoustic time series are often used to investigate rotating machines. This acoustic series are applied together with machine learning approaches to detect anomalies in these machines [Li et al. 2016].

2.2. Acoustic dataset

Compared to other types of data, previous studies have shown that acoustic data—acoustic emissions, in particular—allow for earlier anomaly detection [Saufi et al. 2019]. Despite this, there is a shortage of acoustic datasets, which has motivated the development of solutions in this context. This type of dataset has been extremely important for fault detection in industrial machines, optimizing preventive maintenance. Previous works follow this direction by proposing acoustic data datasets. For example, the MIMII dataset provides sound data for four different industrial machines—slide rails, valves, pumps, and fans—operating under normal and anomalous conditions. Additionally, the dataset includes, for each machine, four different IDs representing four distinct models of the same machine [Purohit et al. 2019]. The lack of diversity and details regarding how the data were collected, the environment characteristics, and the experimental setup motivated the creation of other versions of this dataset [Tanabe et al. 2021, Dohi et al. 2022], in which different aspects of the environment and machine conditions were presented, such as heat, noise, operating speed, and machine load. In addition to MIMII and its versions, other works in this direction have been proposed. The ToyADMOS dataset presents acoustic data from simple mechanical systems, aiming to provide systematically controlled data, offering explainability in the understanding of normal and anomalous conditions [Koizumi et al. 2019]. Another work presents the SOUND-BASED DRONE FAULT dataset, which captures data from drones in normal conditions and anomalous conditions caused by propeller failures and motor issues [Yi et al. 2023]. These works demonstrate the creation of acoustic datasets in different contexts and the application of machine learning to these data for anomaly detection in mechanical systems, highlighting the importance of improving the quality, availability, and diversity of this type of data.

2.3. Cooling fans

Cooling fans are widely used in electronic systems as one of the main devices to dissipate the heat generated by machine components, such as processors, graphics cards, and solid-state drivers. These devices combine metallic heat sinks with fans to increase thermal transfer flow, ensuring that the components operate within safe temperature ranges. The study of these devices is of great importance in fields such as electronics, the automotive sector, climate control, and manufacturing, as efficient cooling directly impacts the health condition of machinery components [Al-Hazmi 2020b]. Recently, cooling fans have been the subject of research on efficiency and operational failures. In one study, the effect of vibrational disturbances caused by the introduction of weights on the blades of the cooling fan was analyzed [Scalabrini Sampaio et al. 2019], considering vibrational for failure analysis. In another study, failure analysis was carried out using vibration data caused by holes in the cooling fan blades [Alhazmi et al. 2025]. The mechanical dynamics observed in cooling fans are not too different when we compare them with more complex rotating machines, such as aircraft turbines—cooling fans can be seen as an instance of these complex machines, and we can use them to gain insights that support a broader understanding of rotating machines.

2.4. Data processing

Acoustic data has rich information that cannot be identified when the raw signal data is analyzed, losing information that can help identify patterns in these data.

To address this issue, signal processing strategies are frequently employed in machine learning pipelines. Techniques such as log-Mel spectrogram [Purohit et al. 2019], Mel-Frequency Cepstral Coefficients (MFCC) [Silva et al. 2024], and statistical features [Bortoni and Jaskowiak 2024] of the signal have gained attention due to successful applications and feature extraction capacity. In particular, MFCC has been widely exploited in problems using acoustic data, such as fake speech detection, speaker recognition, and language and dialect recognition tasks [Abdul and Al-Talabani 2022]. It extracts the most significant features from acoustic data, reducing the data dimension of the signal, which will only have relevant information. Usually, state-of-the-art machine learning models don't work well for acoustic data, and using MFCC in their pipelines, we can significantly improve the performance of these models for tasks involving acoustic data.

3. Dataset Construction

To build the dataset, we first describe the data collection protocol (Subsection 3.1). Followed by the vibratory disturbance model used (Subsection 3.2) and the setup for data collection (Subsection 3.3).

3.1. Recording Protocol

The recording protocol was conducted in a quiet laboratory dedicated solely to data collection, with only ambient noise present. Sounds were collected using a sampling rate of 44.1 kHz and a duration of 10 seconds per recording. Each recording was saved in *.wav* format. Each cooling fan was recorded separately.

3.2. Vibratory disturbance model

A model of vibrational disturbances was employed using neodymium magnets as weights, motivated by recent works on the creation of cooling fan datasets with the introduction of weights on the cooling fan blades [Scalabrini Sampaio et al. 2019]. In this model, a magnet is fixed to each blade of the cooling fan using superglue, forming a layer of magnets. This first layer makes it possible for new magnets to be stacked on top of the existing ones, creating new configurations of weight distribution. We arranged the magnets in four different ways. Each of these arrangements represents an anomaly configuration. The anomaly configurations were labeled as follows: *Config 2, Config 3, Config 4, and Config 5*. In *Config 1*, there are no extra magnets attached to the cooling fan blades—only the first layer, with one magnet on each propeller—so it is considered the normal configuration. Although the anomaly configurations are physically different from each other, we performed a signal analysis to understand the nuances of the signal that we captured and the differences between the signal from normal and abnormal configurations. For this purpose, we selected 40 random normal files and 40 random abnormal files, and then we calculated the FFT difference between random pairs of normal and abnormal files. Finally, we applied the bootstrapping technique to get a 95% confidence interval from the 40 FFT differences. Our finding is shown in Figure 2. Figure 1 illustrates all configurations.

3.3. Data Collection Setup

The equipment listed in Table 1 was selected to support the sound collection in a controlled environment. The combination of signal acquisition, actuation, and structural elements allows the simulation of realistic mechanical behaviors. *Cooling fans* were chosen

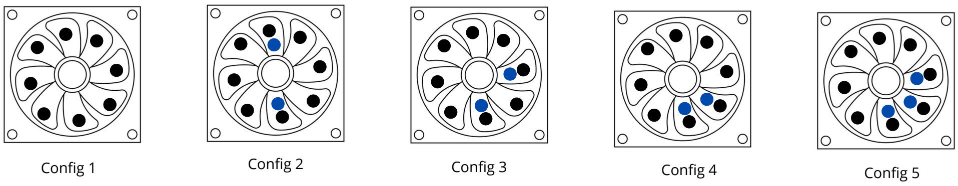

Figure 1. Vibratory disturbance model proposed by [Scalabrini Sampaio et al. 2019]. The black dots represent first-layer magnets attached to the cooling fan blades, while the blue dots indicate additional magnets stacked on top. It is possible to design both balanced configurations (i.e., Config 1) and severely unbalanced ones (i.e., Config 5).

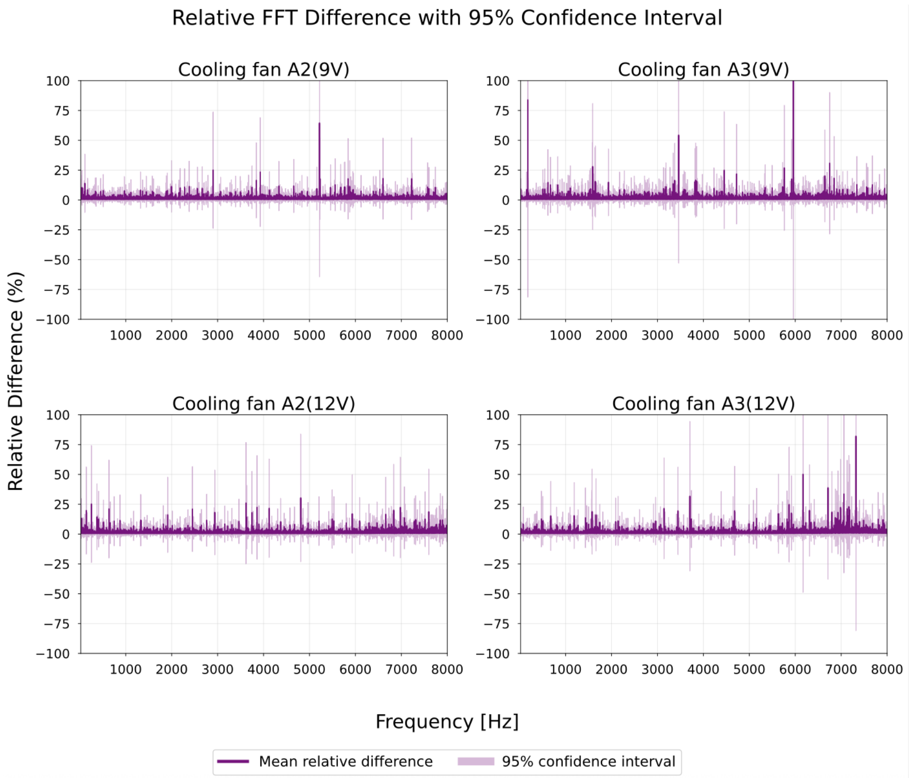

Figure 2. Relative differences in the magnitudes of normal and abnormal data, compared to the magnitude of normal data. Although there are some overlaps, each pair of cooling fan configurations and voltages is distinctly different when comparing the anomalous configurations (configs 2, 3, 4, and 5) to the normal configuration (config 1).

as the main device for data collection, as they are widely used in electronic components for temperature regulation. Magnets were used to introduce configurable perturbations. The sensor used for data collection was a condenser microphone with an amplifier to enhance the capture of the data. The collection platform ensured stability and repeatability throughout the experiments. In Figure 3, we show the complete setup to record all the data.

Table 1. Equipment used in the experimental setup.

Equipment	Model	Amount
Raspberry Pi	Model B	1
Microphone	Audio-Technica AT2020	1
External Source	Icel PS-3005	1
Amplifier	Focusrite Scarlett Solo	1
Cooling Fan	Delta AUC0912D	1
Cooling Fan	Intel AUC0912D	1
Collection Platform	6mm MDF Support Structure	1
Magnet	Neodymium 12x1 mm	100

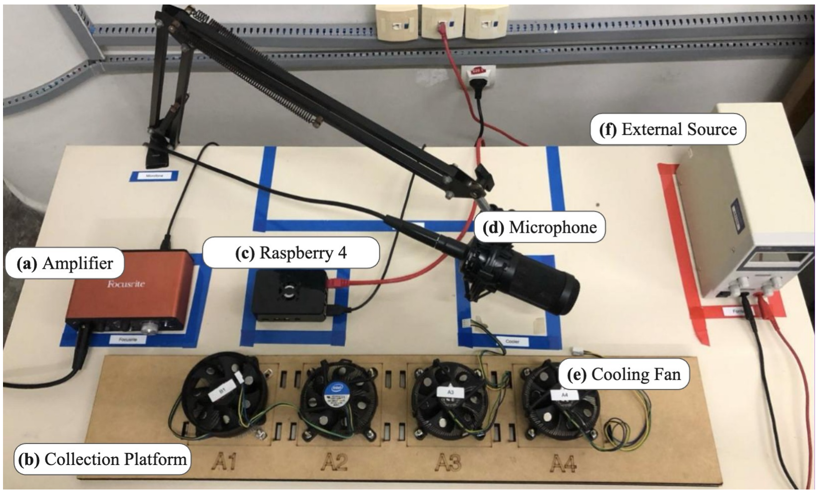

Figure 3. Controlled sound acquisition from cooling fans. The platform (b) includes four labeled slots (A1-A4), each housing a different cooling fan (e). A microphone (d) captures audio signals, which are routed through an amplifier (a) and processed by a Raspberry Pi 4 (c). An external power source (f) provides a stable voltage.

4. Experimental

4.1. Setup

The experimental setup was carefully designed to collect sounds from cooling fans under different weight configurations and voltage levels. To compose our experimental setup, we selected two models of cooling fans: the *Delta AUC0912D*, referred to as *A2*, and the *Intel AUC0912D*, referred to as *A3*—both commonly found in electronic devices. In order to obtain a greater variety of sounds, an experimental arrangement was defined in which five weight configurations were applied to each cooling fan, under two voltage levels: 9

Volts and 12 Volts. For sound processing and storage in *.wav* files, we used a Raspberry Pi 4 and a microphone connected to it for data acquisition. For better reproducibility and manipulation of the machine learning approaches in our proposed dataset, we employed the MTSA framework in our experimental setup. The framework has models ranging from simple to complex, and it allows us to easily compare and do experiments with them in both scenarios: in-distribution (ID) and out-of-distribution (OoD). We prioritized the most explainable and simple models within the framework. The models that we used are Isolation Forest [Liu et al. 2008], OSVM [Schölkopf et al. 2001], and Hitachi [Purohit et al.]. To extract the most important features of the signal, we employed the MFCC feature extraction strategy in the pipeline of all models implemented in the framework.

4.2. Training and Testing

For the training and testing data split, all the anomalous segments were reserved as the test dataset, an equal number of normal segments was randomly selected and reserved as the test dataset, and all the rest of the normal segments were reserved as the training dataset. The training and testing were carried out using the cross-validation strategy, in which parts of the training data are selected individually, and for each of these parts, an instance of the model M_i with corresponding parameters θ_i is trained and evaluated using the AUC-ROC metric with the same test data. The AUC-ROC relates true positive rate and false positive rate, and the area under the curve measures how well the model is classifying true positives compared to false positives. This metric was selected to capture the overall performance of the models.

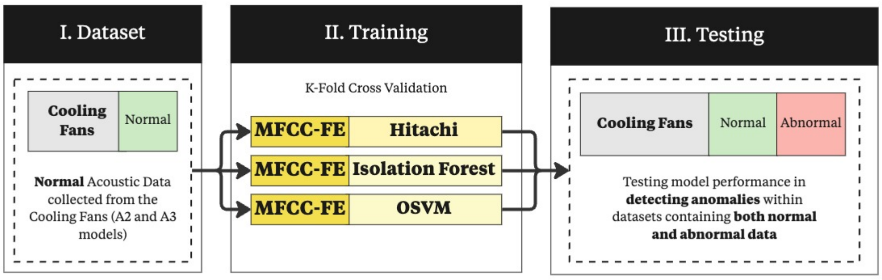

Figure 4. The unsupervised pipeline employed uses cooling fan data, along with machine learning models and features implemented in the MTSA framework.

4.3. Out-of-Distribution Evaluation

For a comprehensive performance analysis, we evaluated the models in both in-distribution (ID) and out-of-distribution (OoD) settings. The ID evaluation involved test data drawn from the same or a very similar distribution as the training data. In this case, training and testing were performed on the same cooling fan model. For the OoD evaluation, we trained the model on A2 and tested it on A3, which comes from a different distribution.

5. Results

All results accomplished by the machine learning approaches with the CANDAS dataset are presented in Table 2. At the end of the data collection process, a total of 10,200 *.wav* files were obtained, which corresponds to approximately 28.33 hours of recording.

Table 2. 95% Confidence Intervals for the AUC-ROC of models trained on Config 1. For In-Distribution, models were trained and tested in the cooling fan type. For Out-of-Distribution, models were trained on A2 but tested on A3.

Config	Volts	Model	ID-(A2, A2)	ID-(A3, A3)	OOD-(A2, A3)
Config 2	9V	Hitachi	0.91 (0.91, 0.91)	0.86 (0.86, 0.86)	0.81 (0.81, 0.82)
		Isolation Forest	0.85 (0.84, 0.85)	0.76 (0.73, 0.79)	0.59 (0.53, 0.64)
		One-Class SVM	0.86 (0.86, 0.86)	0.79 (0.79, 0.80)	0.50 (0.50, 0.50)
	12V	Hitachi	1.00 (1.00, 1.00)	0.97 (0.97, 0.98)	0.97 (0.97, 0.97)
		Isolation Forest	0.92 (0.92, 0.93)	0.84 (0.82, 0.86)	0.60 (0.57, 0.64)
		One-Class SVM	0.91 (0.90, 0.91)	0.89 (0.88, 0.90)	0.50 (0.50, 0.50)
Config 3	9V	Hitachi	0.91 (0.90, 0.91)	0.96 (0.96, 0.96)	0.81 (0.80, 0.81)
		Isolation Forest	0.80 (0.79, 0.81)	0.87 (0.86, 0.87)	0.50 (0.43, 0.55)
		One-Class SVM	0.86 (0.86, 0.86)	0.90 (0.90, 0.91)	0.50 (0.50, 0.50)
	12V	Hitachi	0.96 (0.96, 0.96)	1.00 (1.00, 1.00)	0.96 (0.96, 0.97)
		Isolation Forest	0.94 (0.94, 0.95)	0.94 (0.94, 0.95)	0.77 (0.73, 0.79)
		One-Class SVM	0.89 (0.88, 0.90)	0.89 (0.88, 0.89)	0.50 (0.50, 0.50)
Config 4	9V	Hitachi	0.98 (0.98, 0.98)	0.94 (0.94, 0.94)	0.79 (0.79, 0.80)
		Isolation Forest	0.86 (0.85, 0.86)	0.88 (0.88, 0.89)	0.61 (0.55, 0.67)
		One-Class SVM	0.88 (0.88, 0.89)	0.89 (0.88, 0.89)	0.50 (0.50, 0.50)
	12V	Hitachi	0.99 (0.99, 0.99)	1.00 (1.00, 1.00)	0.89 (0.82, 0.97)
		Isolation Forest	0.96 (0.96, 0.96)	0.95 (0.94, 0.95)	0.74 (0.70, 0.78)
		One-Class SVM	0.89 (0.88, 0.90)	0.89 (0.88, 0.89)	0.50 (0.50, 0.50)
Config 5	9V	Hitachi	0.95 (0.95, 0.95)	0.97 (0.96, 0.97)	0.92 (0.92, 0.92)
		Isolation Forest	0.87 (0.87, 0.88)	0.87 (0.86, 0.88)	0.56 (0.54, 0.59)
		One-Class SVM	0.88 (0.88, 0.89)	0.90 (0.90, 0.91)	0.50 (0.50, 0.50)
	12V	Hitachi	1.00 (1.00, 1.00)	1.00 (1.00, 1.00)	0.53 (0.51, 0.57)
		Isolation Forest	0.96 (0.96, 0.96)	0.96 (0.96, 0.97)	**0.86 (0.84, 0.88)**
		One-Class SVM	0.89 (0.88, 0.90)	0.88 (0.88, 0.89)	0.50 (0.50, 0.50)

5.1. In-Distribution Evaluation

Figure 5 shows the performance of all models, with Hitachi showing the best overall result. When we examine the simpler models—specifically, the One-Class Support Vector Machine (OSVM) and Isolation Forest—we can observe better results for some specific cooling fan models and voltage. The isolation forest model seems more interesting for the cooling fan A2 at 12 volts, accomplishing AUC-ROC values greater than 92 with a very low variance. The same happens with the cooling fan A3 at 12 volts, in which the Isolation Forest accomplishes superior results than OSVM. However, the OSVM model shows superior performance across all cooling fans operating at 9 volts in comparison to Isolation Forest. These results suggest that the most effective model can vary depending on the specific scenario.

5.2. Out-of-Distribution Evaluation by Machines

The results presented in Figure 6 show the performance of all models in the Out-of-distributions scenario. It allows us to analyze the robustness of the model against changes in the distribution of the data when we compare it with the training data. In an overall perspective, the Hitachi model has the best AUC-ROC values. Conversely, the OSVM model is on the random line of all OoD results, representing the worst performance in all OoD scenarios. Remarkably, despite not performing so well in the OoD 9-volt scenario, the Isolation Forest shows a better performance in configuration 5 in the OoD 12-volt scenario, even when compared with the autoencoder Hitachi. These results highlight that in some scenarios, simpler models can be more robust than more complex models, being more resistant to shifts in the data. The Hitachi, along with the other models, showed the smallest difference in AUC ROC between the OoD and ID scenarios.

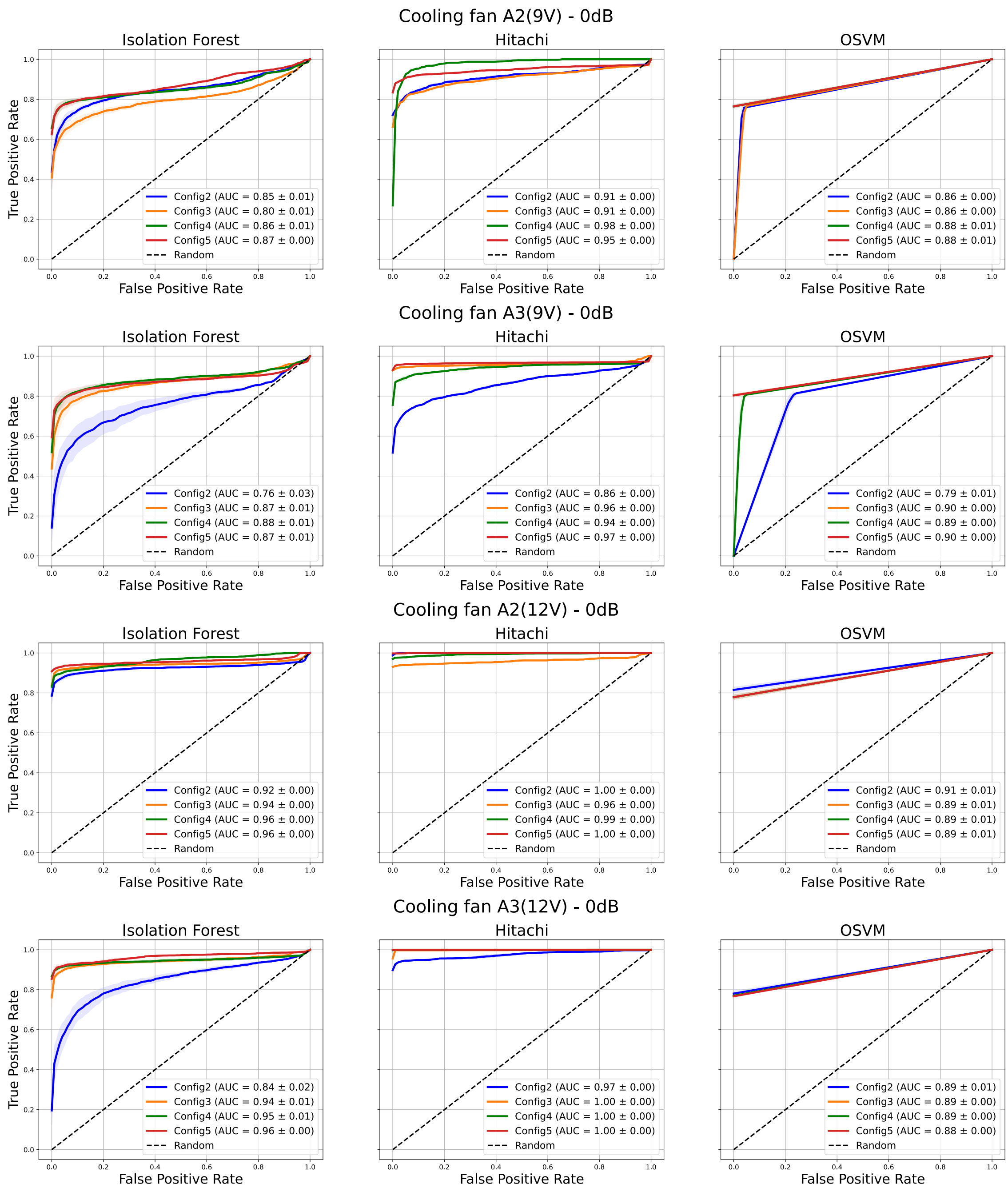

Figure 5. The AUC-ROC analysis of models trained and evaluated on Config1, using the same cooling fan type for both training and testing.

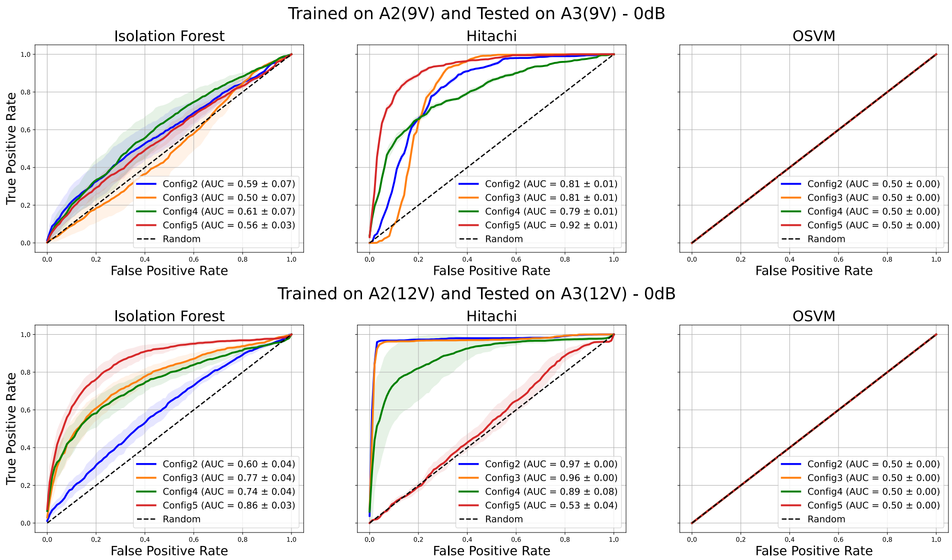

Figure 6. The AUC-ROC analysis of models trained and evaluated on Config 1, in an out-of-distribution scenario where models were trained on A2 and tested on A3.

6. Conclusion and Future Works

This study introduced CANDAS [Lima et al. 2025], a novel sound dataset featuring modeled disturbances and controlled experimental conditions using sounds of cooling fans. The dataset enables the evaluation of different cooling fan types and a variety of anomaly configurations, allowing for a systematic analysis of model performance across diverse scenarios. Our results suggest that, in specific contexts, simpler machine learning models can perform comparably to more complex ones. In addition, by capturing data at different voltage levels, CANDAS provides the opportunity to assess how model performance varies when applied to the same cooling fan type under different voltages. These devices, due to their structural simplicity and widespread presence in computing and industrial environments, serve as an effective pilot platform for accelerating the development of machine learning models aimed at predictive maintenance.

Our findings, obtained through multiple anomaly configurations and a range of machine learning models, demonstrate that even relatively simple approaches can achieve high performance. Furthermore, the out-of-distribution evaluations highlight the dataset's potential for testing the generalization capability of anomaly detection models across different operating conditions. By publicly releasing the CANDAS dataset, we support the machine learning and signal processing research communities in designing and benchmarking robust models. The use of cooling fans as a testbed offers a cost-effective and reproducible strategy to foster a broader understanding of rotating machines and technological maturity in predictive maintenance solutions.

Future work will focus on three main research directions. First, we intend to expand experimental validations to include other types of rotating machinery—such as pumps, turbines, and compressors—to improve the reliability and broaden the applicability of the models to real-world industrial equipment. Second, we plan to enhance our recording setup and equipment, and to explore additional sensors—such as acoustic emission sensors—alongside various machine learning models to develop predictive maintenance pipelines that are both fast and cost-effective. Lastly, we aim to investigate the use of deep learning architectures tailored for fault diagnosis tasks, combining robustness with explainability to enhance model interpretability and practical deployment.

Acknowledgements

The authors acknowledge the financial support by the STIC-AmSud program (STIC AmSud Nro. 23-STIC-06) and by the Coordenacao de Aperfeicoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

References

Abdul, Z. K. and Al-Talabani, A. K. (2022). Mel frequency cepstral coefficient and its applications: A review. *IEEE Access*, 10:122136–122158.

Al-Hazmi, M. W. (2020a). Cooling fan fault diagnostics using vibrational and acoustical analyses. *International Journal of Advanced Science and Technology*, 29(3).

Al-Hazmi, M. W. (2020b). Cooling Fan Fault Diagnostics Using Vibrational and Acoustical Analyses. *International Journal of Advanced Science and Technology*, 29(03).

Alhazmi, M. W., Alsoufi, M. S., Bawazeer, S. A., Hijji, H. H., Alhazmi, H., and Alqurashi, H. F. (2025). Vibration and aeroacoustic analysis of defective cooling fans: effects of blade faults on noise, turbulence, and performance efficiency. *Vibration and aeroacoustic analysis of defective cooling fans: effects of blade faults on noise, turbulence, and performance efficiency*, pages 1–16.

Bortoni, L. and Jaskowiak, P. (2024). Acoustic features and autoencoders for fault detection in rotating machines: A case study. In *Anais da XXXIV Brazilian Conference on Intelligent Systems*, pages 34–49, Porto Alegre, RS, Brasil. SBC.

Chandola, V., Banerjee, A., and Kumar, V. (2009). Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3):1–58.

Chinniah, Y. (2015). Analysis and prevention of serious and fatal accidents related to moving parts of machinery. *Safety science*, 75:163–173.

Dohi, K., Nishida, T., Purohit, H., Tanabe, R., Endo, T., Yamamoto, M., Nikaido, Y., and Kawaguchi, Y. (2022). Mimii dg: Sound dataset for malfunctioning industrial machine investigation and inspection for domain generalization task.

Koizumi, Y., Saito, S., Uematsu, H., Harada, N., and Imoto, K. (2019). Toyadmos: A dataset of miniature-machine operating sounds for anomalous sound detection.

Li, C., Sanchez, R.-V., Zurita, G., Cerrada, M., Cabrera, D., and Vásquez, R. E. (2016). Gearbox fault diagnosis based on deep random forest fusion of acoustic and vibratory signals. *Mechanical Systems and Signal Processing*, 76-77:283–293.

Li, Y., Yu, R., Shahabi, C., and Liu, Y. (2018). Diffusion convolutional recurrent neural network: Data-driven traffic forecasting.

Lima, H., Silva, C., Nery, R., Camacho, W. C., Brazileiro, R., de Paula Monteiro, R., Ribeiro, A. M. N. C., Bastos-Filho, C. J. A., Cerrada, M., and Pinheiro, D. (2025). IoTDataAtelier/CANDAS · Datasets at Hugging Face — huggingface.co. `https://huggingface.co/datasets/IoTDataAtelier/CANDAS`.

Liu, F. T., Ting, K. M., and Zhou, Z.-H. (2008). Isolation forest. In *2008 Eighth IEEE International Conference on Data Mining*, pages 413–422. IEEE.

Moura Filho, L. F., de Paula Monteiro, R., Pinheiro, D., Endo, P. T., and Ribeiro, A. M. N. (2023). Forecasting imminent failures in electrical industrial centrifuge using machine learning. In *2023 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*, pages 1–6. IEEE.

Purohit, H., Tanabe, R., Ichige, K., Endo, T., Nikaido, Y., Suefusa, K., and Kawaguchi, Y. MIMII dataset: Sound dataset for malfunctioning industrial machine investigation and inspection. Number: arXiv:1909.09347.

Purohit, H., Tanabe, R., Ichige, K., Endo, T., Nikaido, Y., Suefusa, K., and Kawaguchi, Y. (2019). Mimii dataset: Sound dataset for malfunctioning industrial machine investigation and inspection.

Roelofs, C. M., Gück, C., and Faulstich, S. (2024). Transfer learning applications for autoencoder-based anomaly detection in wind turbines. *Energy and AI*, 17:100373.

Roy, R. and George, K. T. (2017). Detecting insurance claims fraud using machine learning techniques. In *2017 International Conference on Circuit ,Power and Computing Technologies (ICCPCT)*, pages 1–6.

Saufi, S. R., Ahmad, Z. A. B., Leong, M. S., and Lim, M. H. (2019). Challenges and Opportunities of Deep Learning Models for Machinery Fault Detection and Diagnosis: A Review. *IEEE Access*, 7:122644–122662.

Scalabrini Sampaio, G., Vallim Filho, A. R. D. A., Santos Da Silva, L., and Augusto Da Silva, L. (2019). Prediction of Motor Failure Time Using An Artificial Neural Network. *Sensors*, 19(19):4342.

Schölkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J., and Williamson, R. C. (2001). Estimating the support of a high-dimensional distribution. *Neural Computation*, 13(7):1443–1471.

Silva, C., Coelho, M., Barros, U., Ribeiro, A. M. N. C., de Paula Monteiro, R., and Pinheiro, D. (2024). A unified framework for replicability of anomaly detection in multiple time series: Enhancing ganf and ransyncoders pipelines with mfcc feature extraction for acoustic data. In *2024 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*, pages 1–6.

Tanabe, R., Purohit, H., Dohi, K., Endo, T., Nikaido, Y., Nakamura, T., and Kawaguchi, Y. (2021). Mimii due: Sound dataset for malfunctioning industrial machine investigation and inspection with domain shifts due to changes in operational and environmental conditions.

Yi, W., Choi, J.-W., and Lee, J.-W. (2023). Sound-based drone fault classification using multitask learning.

Zhou, H., Yu, K., Zhang, X., Wu, G., and Yazidi, A. (2022). Contrastive autoencoder for anomaly detection in multivariate time series. *Information Sciences*, 610:266–280.