Analysis of time–frequency signal representations for anomaly detection in industrial environments

Análise de representações tempo–frequência de sinais para detecção de anomalias em ambientes industriais

Pedro Henrique Meira de Araújo[1]
orcid.org/0000-0003-2600-3886

Rodrigo de Paula Monteiro[2]
orcid.org/0000-0002-2423-5088

Carmelo José Albanez Bastos Filho[3]
orcid.org/0000-0002-0924-5341

[1]Escola Politécnica de Pernambuco, Universidade de Pernambuco, Recife, Brasil.
E-mail: phma@ecomp.poli.br

[2]Escola Politécnica de Pernambuco, Universidade de Pernambuco, Recife, Brasil. Email: rodrigo.monteiro@poli.br

[3]Escola Politécnica de Pernambuco, Universidade de Pernambuco, Recife, Brasil. Email: carmelo.filho@upe.br

DOI: 10.25286/repa.v11i3.3904

REVISTA DE Engenharia e Pesquisa Aplicada

RESUMO

No contexto da Indústria 4.0, a detecção de anomalias baseada em som tem se consolidado como uma estratégia relevante para apoiar a manutenção preditiva e reduzir paradas não programadas em ambientes industriais. Este estudo investiga o impacto de diferentes representações tempo–frequência na detecção de anomalias acústicas em máquinas industriais operando sob condições ruidosas. Utilizando a base de dados MIMII, foram conduzidos experimentos com dois elementos de máquinas (bomba e deslizante) sob três níveis de relação sinal-ruído (SNR): 6 dB, 0 dB e −6 dB. Três representações de características, espectrograma, MelSpectrograma e coeficientes cepstrais em frequência Mel (MFCC), foram geradas e utilizadas como entrada de uma rede neural convolucional para classificação binária entre estados normais e anômalos. Além das métricas clássicas de avaliação, como acurácia, precisão e recall, mapas de divergência de Kullback–Leibler foram empregados para analisar a separabilidade entre padrões acústicos. Os resultados indicam que o MelSpectrograma apresenta desempenho mais estável e consistente entre os diferentes níveis de SNR, especialmente sob ruído severo, enquanto o MFCC se mostra mais sensível às variações de ruído. Testes estatísticos baseados no método de Wilcoxon confirmam a maior robustez do MelSpectrograma em comparação com as demais representações, destacando seu potencial para aplicações futuras em detecção de anomalias acústicas industriais.

PALAVRAS-CHAVE: *Indústria 4.0; Detecção de Anomalias; Aprendizado de Máquina; Representação Tempo–Frequência.*

ABSTRACT

In the context of Industry 4.0, sound-based anomaly detection has emerged as a relevant strategy to support predictive maintenance and reduce unplanned downtime in industrial environments. This study investigates the impact of different time–frequency representations on acoustic anomaly detection in industrial machines operating under noisy conditions. Using the MIMII dataset, experiments were conducted on two machine elements (pump and slider) under three signal-to-noise ratio (SNR) levels: 6 dB, 0 dB, and −6 dB. Three feature representations, spectrogram, MelSpectrogram, and Mel-frequency cepstral coefficients (MFCC), were generated and used as inputs to a convolutional neural network for binary classification between normal and anomalous states. In addition to standard evaluation metrics such as accuracy, precision, and recall, Kullback–Leibler divergence maps were employed to analyze the separability between acoustic patterns. The results indicate that MelSpectrogram provides more stable and consistent performance across different SNR levels, especially under severe noise conditions, while MFCC proves to be more sensitive to noise variations. Statistical tests based on the Wilcoxon method confirm the superior robustness of MelSpectrogram compared to the other representations, highlighting its potential for future applications in industrial acoustic anomaly detection.

KEYWORDS: *Industry 4.0, Anomaly Detection, Machine Learning, Time Frequency Representation*

1 INTRODUCTION

In the context of Industry 4.0, Artificial Intelligence (AI) has been widely adopted in several fields, including computer vision, robotics, and data analytics. These applications constitute a key element of the digital transformation across modern industrial sectors. Within this scenario, predictive maintenance has emerged as an essential strategy for monitoring machine conditions and exploiting data to ensure the reliable and efficient operation of industrial assets [1-2]. Anomaly detection is a central component of such strategies, as it contributes to the reduction of unplanned downtime and the mitigation of associated economic losses. Consequently, Machine Learning (ML) techniques have been increasingly employed to process large volumes of data, identify patterns, and perform classification tasks with improved accuracy [3].

In predictive maintenance applications, ML methods constitute a fundamental tool for enhancing the classification of industrial data [3]. These methods are capable of integrating measurements from heterogeneous sensors and extracting latent structures from the observed signals. The choice and representation of sensor data are critical factors that influence the performance of anomaly classification in industrial machinery. In particular, robust signal representations are required in noisy environments, where suitable feature extraction can significantly improve the reliability of anomaly detection systems [4].

Several studies, such as [5-7], have employed acoustic datasets for fault classification in rotating or stationary machinery. However, these works do not provide a systematic comparison of different signal representation schemes. Other contributions have focused on comparative analyses of time–frequency representations. For example, some studies compare multiple time–frequency techniques for condition monitoring [8], whereas our previous work [9] assess classification performance under varying signal-to-noise ratio (SNR) levels, often restricting the analysis to a single type of component or operating condition.

ML algorithms can operate on various forms of input data; nonetheless, time–frequency representations are often preferred due to the rich descriptive content they provide to neural network models. Despite their extensive use in the literature, there remains a gap in understanding how different representation techniques interact with changing noise conditions and how these interactions affect classification performance. Different time–frequency representations can markedly influence the ability of a model to discriminate between normal and anomalous operating states, particularly in environments characterized by significant noise.

Therefore, this study seeks to evaluate the impact of multiple time–frequency representation methods on anomaly classification performance over a range of SNR scenarios and to elucidate the extent to which each representation contributes to the robustness of ML-based predictive maintenance systems in industrial settings.

2 MATERIALS AND METHODS

2.1 DATASET

We used the Malfunctioning Industrial Machine Investigation and Inspection (MIMII) dataset, a sound dataset that consists of files in the format .wav, referring to the states normal and abnormal of machine elements, such as pump, valve, slider, and fan. These devices are elements used in industrial environments and their behavior influences directly in the machine's operation. Understanding the normals and abnormal patterns of these elements can help to classify anomalies and provide support to decision for maintenance managers.

In the slide rail case, the anomalous conditions include rail damage, a loose drive belt, and lack of grease in the moving parts [10]. These defects change the friction and the dynamics of the linear motion, generating acoustic signatures that differ from those observed under normal operation.

In the pump case, anomalies mainly involve leakage, contamination, and clogging in the system [10]. These issues alter the fluid flow and the mechanical load on the equipment, which is reflected in changes in the sound characteristics captured in the dataset recordings

The sounds are captured through eight microphones, which can capture 16-bit signals with a sample of 16 KHz[10].

For each machine type, three noise conditions were considered, with signal-to-noise ratio (SNR) values of 6 dB, 0 dB, and −6 dB. The goal of this work is to evaluate how different time–frequency representations perform under these three noise conditions. In particular, two machine types from

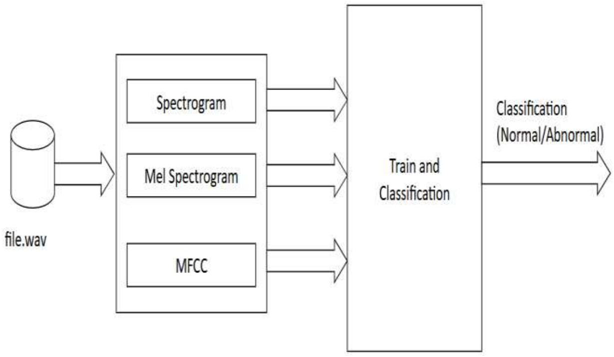
DOI: 10.25286/repa.v11i3.3904

the MIMII dataset were selected to analyze the impact of time–frequency representations across all SNR levels.

2.2 EXPERIMENTAL SETUP AND DATA PROCESSING

The process of generating the time–frequency representations is carried out with each sound of the dataset, like we can see in Figure 1. Firstly, the conversion occurred using the fast fourier transform, which produces a spectrogram representation [7]. The same set of WAV files is employed to derive both the Mel-spectrogram representations [11] and the Mel-frequency cepstral coefficients (MFCCs) [12], using the short-time Fourier transform (STFT) followed by the specific processing steps required for each feature type. These three sets of signal representations are used as input to a convolutional neural network to realize the train and classification. In the final part of the process the system learns and the outputs can show if the inputs are normal or abnormal behavior.

Figure 1 – Proposed Method



Source: Author

2.3 PREPROCESSING EXPERIMENTAL DESIGN

Figure 1 present spectrograms, mel spectrograms and Mel - frequency to generated times - frequency representations aim classifying normal and anomalous instances.

We generated the spectrograms utilizing STF with n_fft= 1024, win_length=1024, hop_length=256 and power =2. This method are applied in all wav files of all equipment in the data set. To produce the mel spectrograms are utilized Mel Spectrograms function from the torchaudio.transforms library. The parameters are n_fft=1024, win_legth=1024, hop_length=256, power =2, n_mels=128 and sample_rate=16000. The same configuration was applied in all files.

To produce the MFCC we utilize the MFCC function, where n_mfcc=40, n_fft=1024, n_mels=128, win_length=1024, hop_length=256 and power=2.0. We applied this set of parameters in every signals.

2.4 AUDIO BASED ANOMALY CLASSIFICATION

To perform the classification, we input the time-frequency signal representations into a Convolutional Neural Network (CNN). The spectrograms were transformed into arrays and used as input data. We trained the CNN with a learning rate of 0.001, employing a cross-entropy loss function and optimizing it with the Adam algorithm.

The batch size used was determined based on the data set characteristics. We trained the CNN over 500 epochs, dividing the data sets into normal and abnormal categories for each type of equipment (pump, valve, slider, and fan). Each training and test set consisted of 25 normal and 25 abnormal samples. Additionally, 20% of the data was set aside for validation.

Table I summarizes the architecture of the convolutional neural network used in this study. The model receives single-channel time–frequency images as input and processes them through three convolutional blocks with pooling, followed by a dropout layer, a fully connected layer, and a final linear layer for binary classification. The models will be run 30 times, and the evaluation will be performed using accuracy, precision, and recall as metrics.

Table I -Convolutional neural network architecture for audio-based anomaly detection

Stage	Layer / Operation	Main parameters / description
Input	Input image	1 × H × W (grayscale TF map)
Convolutional block 1	Conv2d + ReLU + MaxPool2d	16 filters, 3×3, padding = same; max-pool 2×2
Convolutional block 2	Conv2d + ReLU + MaxPool2d	32 filters, 3×3, padding = same; max-pool 2×2
Convolutional block 3	Conv2d + ReLU + MaxPool2d	64 filters, 3×3, padding = same; max-pool 2×2
Regularization	Dropout	p = 0.2 (after conv blocks)
Fully connected	Linear + activation (e.g., ReLU)	64 × 8 × 8 dense_neurons (e.g., 32)
Output	Linear	dense_neurons 2 (binary classes)

Source: Author

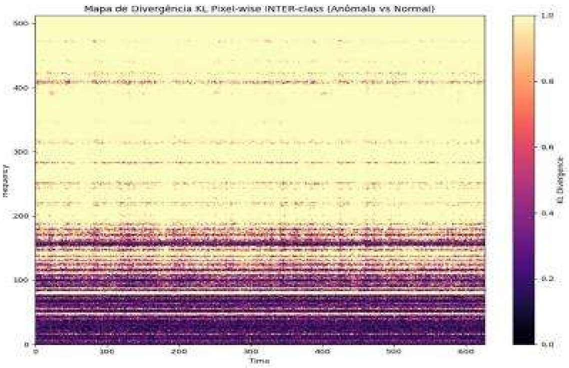
3.RESULTS

In this section, we present the classification results for two machine elements, pump and slider, together with the corresponding evaluation metrics and their behavior. In addition, we include, as an illustrative example, a Kullback–Leibler divergence analysis for one of the studied elements, in order to visually present the divergence between normal and anomalous conditions.

3.1 KULLBACK-LEIBLER DIVERGENCE

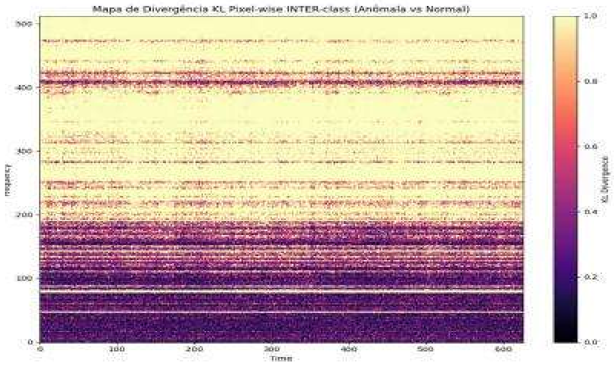
Figures 2, 3, and 4 show the interclass KL divergence maps for 6 dB, 0 dB, and −6 dB to spectrograms originating from the slider [13]. At 6 dB, the map displays clear and structured regions of divergence, indicating strong separability between anomalous and normal signals. As the SNR decreases to 0 dB, these regions become more diffuse, reflecting the impact of noise on the discriminative characteristics of the signals. At −6 dB, the map loses most of its structure, suggesting that the high noise level masks relevant spectral features and substantially reduces class separability. This scenario makes anomaly detection more difficult due to the reduced ability to distinguish between the two classes.

Figure 2 – Interclass Divergence Map for 6 dB



Source: Author

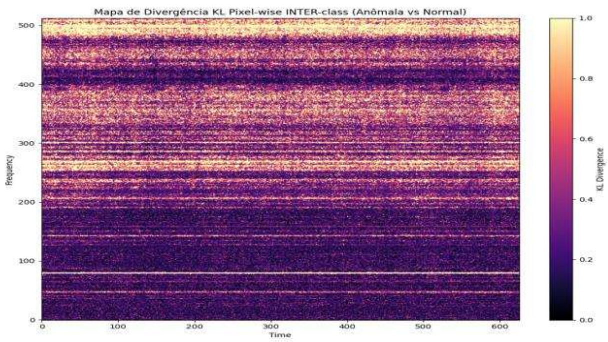
Figure 3 – Interclass Divergence Map for 0dB



Source: Author

DOI: 10.25286/repa.v11i3.3904

Figure 4 – Interclass Divergence Map for -6dB



Source: Author

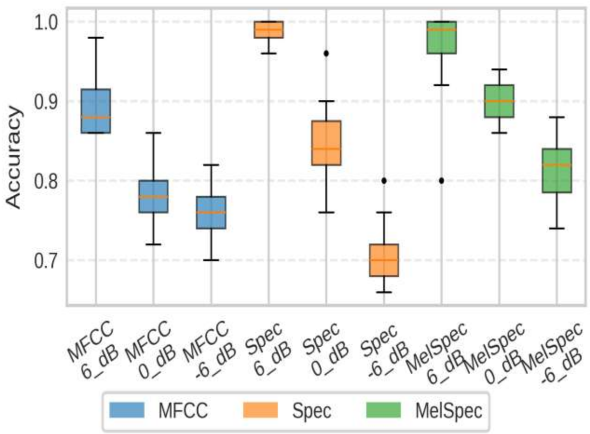
3.2 CLASSIFICATION

The classification results obtained for each machine element are presented and compared in this section.

3.2.1 Slider

According to the boxplots presented in Figure 5, in the scenario with best SNR (Signal noise relationship) the spectrograms and MelSpectrograms achieved the best results of accuracy. With the change of SNR, the melspectrogram shows more constancy of its accuracy.

Figure 5 – SLIDER - Accuracy - Feature Representation and Noise Level



Source: Author

The average precision showed that in two of the three scenarios MelSPec provides the best precision and consistency following for the Spec, which can be observed in Table II.

Table II - Precision - Slider

SNR	6dB	0dB	-6dB
MFCC	0.83 (0.05)	0.77 (0.04)	0.80 (0.07)
Spec	0.98 (0.02)	0.82 (0.05)	0.71 (0.05)
MelSpec	0.99 (0.01)	0.87 (0.04)	0.77 (0.04)

Source: Author

Table III presents the results for the soft difference between MelSpec to other representations in 6dB. However in other SNR scenarios the values of MelSpec are more the MFCC and Spec. The MFCC presented one more difference between 6dB and 0dB.

Table III - Recall - Slider

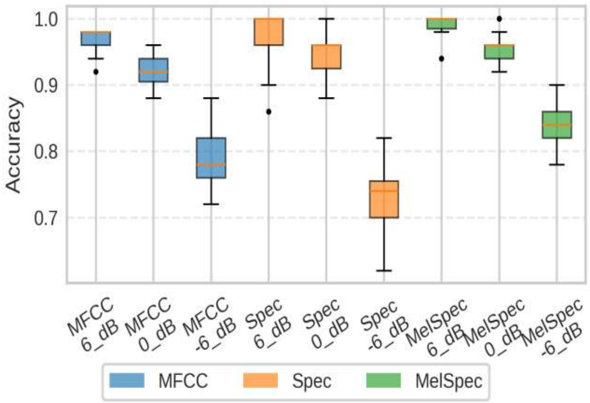
SNR	6dB	0dB	-6dB
MFCC	0.98 (0.03)	0.78 (0.07)	0.78 (0.05)
Spec	0.98 (0.03)	0.87 (0.06)	0.87 (0.07)
MelSpec	0.95 (0.08)	0.93 (0.03)	0.93 (0.06)

Source: Author

3.2.2 Pump

Figure 6 presents the results of the accuracy presented in 6dB as the low variation in MelSpectrogram with the best results followed for the spectrogram. The same behaviorism is presented in 0dB. The lower SNR presented best results with MelSpectrograms, indicating more consistency for the Mel representations.

Figure 6 –PUMP- Accuracy - Feature Representation and Noise Level



Source: Author

Table IV presents the results for precision of the pump and reveals a variation where in each SNR there is a representation of best value. The spec presented the high variation, specially when you can observe the 0db and -6dB values. Table V presents the MelSpec as the lower variation, MFCC has the second highest results.

Table IV - Precision - Pump

SNR	6dB	0dB	-6dB
MFCC	0.95 (0.02)	0.88(0.03)	0.86 (0.04)
Spec	0.99 (0.02)	0.94 (0.05)	0.76 (0.06)
MelSpec	0.99 (0.02)	0.93 (0.03)	0.85 (0.05)

Source: Author

Table V - Recall - Pump

SNR	6dB	0dB	-6dB
MFCC	0.98 (0.02)	0.96 (0.02)	0.69 (0.08)
Spec	0.94 (0.08)	0.96 (0.05)	0.65 (0.08)
MelSpec	0.99 (0.03)	0.98 (0.02)	0.85 (0.05)

Source: Author

The advantage of MelSpec is even greater in relation to slider, especially in Recall, where the difference compared to Spec and MFCC becomes more pronounced at −6 dB.”

In general we can observe that MelSpectrogram presents more consistent results in face of the variations of SNR when compared to the other forms of time-frequency rerepresentations. The spectrogram performs well at high SNR levels, but experiences significant degradation as the noise increases.

3.3 STATISTICAL SIGNIFICANCE AND ROBUSTNESSS

This subsection investigates the statistical significance and robustness of the proposed feature representations under different noise conditions. To this end, Wilcoxon signed-rank tests are applied to compare pairs of representations at fixed SNR levels (Tables V and VII) and to analyze the effect of varying SNR within each representation type (Tables VI and VIII).

3.3.1 Slider

The interclass and intraclass Wilcoxon tests yielded p-values below 0.05 regarding all tested scenarios, suggesting that the compared samples originate from statistically different distributions. These results confirm that MelSpectrogram consistently provides the best performance, as discussed in Table VI. Table VII reports uniformly low p-values, further showing that changes in noise level strongly affect spectrogram-based representations. In contrast, MFCC features exhibited the lowest robustness across the evaluated metrics, which is expected given that MFCCs are primarily tailored to speech rather than industrial acoustic signals.

DOI: 10.25286/repa.v11i3.3904

Table VI - Wilcoxon test results comparing different noise levels across representation types

Types of Representation (Comparison)	Noise Level	P-value
MFCC vs Spectrogram	6dB	0.000087
MFCC vs MelSpectrogram	6dB	0.000003
Spectrogram vs MelSpectrogram	6dB	0.003816
MFCC vs Spectrogram	0dB	0.004906
MFCC vs MelSpectrogram	0dB	0.000001
Spectrogram vs MelSpectrogram	0dB	0.001046
MFCC vs Spectrogram	-6dB	0.000002
MFCC vs MelSpectrogram	-6dB	0.000005
Spectrogram vs MelSpectrogram	-6dB	0.02

Source: Author

Table VII - Wilcoxon test of methods across different noise levels

Types of Representation	Comparison	P-value
MFCC	6dB vs 0dB	0.000003
MFCC	6dB vs -6dB	0.000002
MFCC	0dB vs -6dB	0.000002
Spec	6dB vs 0dB	0.000002
Spec	6dB vs -6dB	0.000001
Spec	0dB vs -6dB	0.001560
MelSpec	6dB vs 0dB	0.000003
MelSpec	6dB vs -6dB	0.000002
MelSpec	0dB vs -6dB	0.00001

Source: Author

3.3.2 Pump

Table VIII presents the Wilcoxon test revealed in that MelSpectrogram is significantly more robust at high and low SNR levels (6 dB and −6 dB), outperforming both MFCC and standard Spectrogram. At moderate noise (0 dB), MelSpectrogram and Spectrogram perform similarly, while MFCC shows significantly inferior stability. Overall, MFCC is the representation most affected by noise, whereas MelSpectrogram consistently provides the strongest discrimination between normal and anomalous acoustic patterns.

Table VIII - Wilcoxon test results comparing different noise levels across representation types

Types of Representation (Comparison)	Noise Level	P-value
MFCC vs Spectrogram	6dB	0.345901
MFCC vs MelSpectrogram	6dB	0.000168
Spectrogram vs MelSpectrogram	6dB	0.009925
MFCC vs Spectrogram	0dB	0.002352
MFCC vs MelSpectrogram	0dB	0.000052
Spectrogram vs MelSpectrogram	0dB	0.806142
MFCC vs Spectrogram	-6dB	0.000024
MFCC vs MelSpectrogram	-6dB	0.000068
Spectrogram vs MelSpectrogram	-6dB	0.000002

Source: Author

Table IX results show that the most suitable feature representation depends on the noise level, with some representations being interchangeable under certain conditions and clearly distinct under others. At 6 dB, MFCC and spectrogram perform equivalently, while MelSpectrogram differs significantly from both. At 0 dB, spectrogram and

MelSpectrogram become statistically indistinguishable, and MFCC departs from them. At −6 dB, all pairwise comparisons among MFCC, spectrogram, and MelSpectrogram are statistically different. Thus, under severe noise, the choice of representation substantially affects performance.

Table IX - Wilcoxon test of methods across different noise levels

Types of Representation	Comparison	P-value
MFCC	6dB vs 0dB	0.000002
MFCC	6dB vs -6dB	0.000002
MFCC	0dB vs -6dB	0.000002
Spec	6dB vs 0dB	0.006706
Spec	6dB vs -6dB	0.000002
Spec	0dB vs -6dB	0.000002
MelSpec	6dB vs 0dB	0.000011
MelSpec	6dB vs -6dB	0.000002
MelSpec	0dB vs -6dB	0.000002

Source: Author

4 CONCLUSIONS

In this study, different time–frequency representations were systematically evaluated for audio-based anomaly detection in multiple industrial machine elements (e.g., pump and slider) under varying signal-to-noise ratios. The results show that MelSpectrogram consistently provides more stable and accurate performance across SNR levels, particularly at −6 dB, where it maintains higher precision and recall than MFCC and standard spectrograms. Although spectrograms achieve competitive results at higher SNRs, their performance degrades more sharply as noise increases, while MFCC is the representation most affected by SNR variations. These findings suggest that MelSpectrogram is a suitable default choice for robust anomaly detection in audio datasets derived from industrial equipment, and it will therefore be adopted as the time–frequency representation method in a multisensor anomaly detection pipeline currently under development.

5 REFERENCES

[1] SERRADILLA, Oscar; ZUGASTI, Ekhi; RODRIGUEZ, Jon; ZURU TUZA, Urko. Deep learning models for predictive maintenance: a survey, comparison, challenges and prospects. **Applied Intelligence**, v. 52, n. 10, p. 10934–10964, 2022. [2] D. Velasquez, E. Perez, X. Oregui, A. Artetxe, J. Manteca, J. E. Mansilla, M. Toro, M. Maiza, and B. Sierra, A hybrid machine-learning ensemble for anomaly detection in real-time Industry 4.0 systems, IEEE Access, vol. 10, pp. 72024–72036, 2022.

[2] D. Velasquez, E. Perez, X. Oregui, A. Artetxe, J. Manteca, J. E. Mansilla, M. Toro, M. Maiza, and B. Sierra, A hybrid machine-learning ensemble for anomaly detection in real-time Industry 4.0 systems, **IEEE Access**, vol. 10, pp. 72024–72036, 2022.

[3] T. Ye, T. Peng, and L. Yang, Review on sound-based industrial predictive maintenance: From feature engineering to deep learning, **Mathematics**, vol. 13, no. 11, p. 1724, 2025.

[4] LISO, Adriano; CARDELLICCHIO, Angelo; PATRUNO, Cosimo; NITTI, Massimiliano; ARDINO, Pierfrancesco; STELLA, Ettore; REN` O, Vito. A review of deep learning based anomaly detection strate gies in Industry 4.0 focused on application fields, sensing equipment and algorithms. **IEEE Access**, 2024.

[5] DE OLIVEIRA NETO, Wilson A.; GUEDES, Ello´a B.; FIGUEIREDO, Carlos Maur´ ıcio S. Anomaly Detection in Sound Activity with Gen erative Adversarial Network Models. Journal of Internet Services and Applications, v. 15, n. 1, p. 313–324, 2024.**[6]** MERANEH, Awaleh Houssein; AUTREL, Fabien; BOUDER, MERANEH, Awaleh Houssein; AUTREL, Fabien; BOUDER, H´el`ene Le; PAHL, Marc-Oliver. SADIS: real-time sound-based anomaly detec tion for industrial systems. **In: International Symposium on Foundations and Practice of Security**. Springer, 2023. p. 82–92.

[7] H. Lee and J. Yu, A fault detection framework for rotating machinery with a spectrogram and convolutional autoencoder, **Applied Sciences**, vol. 15, no. 14, p. 7698, 2025. doi: 10.3390/app15147698.

[8] WANG, Mei; MEI, Qingshan; SONG, Xiyu; LIU, Xin; KAN, Ruix iang; YAO, Fangzhi; XIONG, Junhan; QIU, Hongbing. A machine anomalous

DOI: 10.25286/repa.v11i3.3904

sound detection method using the LMS spectrogram and ES MobileNetV3 network. **Applied Sciences**, v. 13, n. 23, p. 12912, 2023

[9] P. H. M. Araújo; R. P. Monteiro; R. B. B. HOLANDA; C. J. A. B. Filho; M.C. Lozad**a**. Comparative Study of Time-Frequency Representations for Anomaly Detection in Industrial Equipment. **In: XVII Congresso Brasileiro de Inteligência Computacional (CBIC)**, 2025, Belo Horizonte. In press.

[10] H. Purohit, Y. Koizumi, S. Saito, and N. Harada, MIMII Dataset: Sound Dataset for Malfunctioning Industrial Machine Investigation, Zenodo, 2019. Available: https://zenodo.org/record/3384388

[11] E. Yun and M. Jeong, Acoustic feature extraction and classification techniques for anomaly sound detection in the electronic motor of automotive EPS, **IEEE Access**, 2024.

[12] Peng Jiang, Yuhui Wang, Shuang Wu, Luying Zhang, and Chang Yang. Fault diagnosis of wind turbine pitch bearings via transfer learning and an improved residual network. **Renewable Energy**, Elsevier, 2022.