Comparative Study of Time–Frequency Representations for Anomaly Detection in Industrial Equipment

Pedro Henrique Meira de Araújo
Escola Politécnica de Pernambuco
Universidade de Pernambuco
Recife, Brazil
phma@ecomp.poli.br

Rodrigo de Paula Monteiro
Escola Politécnica de Pernambuco
Universidade de Pernambuco
Recife, Brazil
rodrigo.monteiro@poli.br

Rafael Bérgamo Barreto de Holanda
Escola Politécnica de Pernambuco
Universidade de Pernambuco
Recife, Brazil
rafael.holanda@upe.br

Carmelo José Albanez Bastos Filho
Escola Politécnica de Pernambuco
Universidade de Pernambuco
Recife, Brazil
carmelo.filho@upe.br

Mariela Cerrada Lozada
Facultad de Investigación
Universidad Estatal de Milagro
Milagro, Ecuador
mcerradal@unemi.edu.ec

***Abstract*—In Industry 4.0, companies focus on predictive maintenance to reduce expenses and limit asset depreciation. We actively investigate anomaly detection as a way to identify potential equipment failures. Given the critical role of rotating machines in various industrial applications, our study examines how images generated from the signals of rotating machine components can aid in modeling a classification network for detecting anomalies. We utilized the MIMII dataset and implemented classification techniques using Spectrogram, Mel Spectrogram, and MFCCs methods derived from both anomalous and normal signals. We employed the generated images in a Convolutional Neural Network to distinguish between anomalous and normal instances. The Mel spectrogram clearly outshined the other methods, particularly when faced with fluctuations in signal noise. At an SNR of 6 dB, it achieved remarkable accuracy and recall scores of 0.99 and 0.98, respectively. Even when the conditions worsened to -6 dB, the Mel spectrogram still delivered impressive results, boasting an accuracy of 0.87 and a recall of 0.90, which underscores its remarkable stability and performance under challenging conditions. In comparison, the traditional Spectrogram recorded both accuracy and recall at 0.85 in the same -6 dB scenario, while the MFCC Spectrogram lagged behind significantly, with scores of only 0.74 for accuracy and 0.66 for recall. These findings highlight the Mel spectrogram's superior effectiveness in tackling noisy environments. This work sets the stage for developing a framework that enhances machine learning techniques for detecting anomalies in industrial machinery.**

***Index Terms*—industry 4.0, predictive maintenance, rotating machines, machine learning, anomaly detection**

I. Introduction

Predictive maintenance has emerged as an essential approach in industrial operations, aiming to minimize unplanned downtime, extend equipment lifespan, and optimize maintenance schedules. This approach is closely related to anomaly detection, which focuses on identifying deviations from normal operational behavior that may indicate potential failures. Early detection of such anomalies is imperative for preventing costly interruptions and ensuring system reliability [1].

In recent years, machine learning (ML) techniques have become important tools for predictive maintenance and anomaly detection. These methods enable the analysis of vast amounts of sensor data to uncover patterns and identify irregularities that traditional rule-based systems might miss [2]. In this context, deep learning models have demonstrated significant performance improvement in capturing complex relationships within data, leading to more accurate and timely detection of equipment anomalies [3].

A critical factor influencing the performance of ML models in anomaly detection is the representation of input data. The way raw sensor data, such as audio signals, is transformed into features can significantly impact a model's ability to distinguish between normal and anomalous behavior. Effective data representation is crucial for improving model accuracy and robustness, particularly in noisy environments.

Although deep learning models are increasingly adopted for anomaly detection in audio-based predictive maintenance, there is a lack of studies comparing different data representation techniques under varying noise conditions [4]. Many existing research works on anomaly detection, such as the ones developed by Neto *et al.* [5] and Meraneh *et al.* [6], focus on the anomaly detection models and do not perform a comparative analysis between different representation techniques [7]. Other studies, like the one developed by Wang et al. [8], explore representation techniques but do not account for the impact of varying noise levels. This gap becomes particularly relevant in industrial environments, where background noise is often unavoidable and can significantly interfere with the

extraction of meaningful signal features [9].

This study addresses a gap in the field by conducting a comparative analysis of three commonly used audio representations: Spectrogram, Mel Spectrogram, and Mel Frequency Cepstral Coefficients (MFCCs). We evaluated the effectiveness of these representations for anomaly detection using the MIMII dataset [9], which includes audio recordings of industrial machinery operating under both normal and faulty conditions, with varying levels of background noise. We input each audio representation into a classification model and analyzed the results obtained from each one. Our objective is to assess the robustness and discriminative power of each representation in noisy environments, ultimately contributing to the development of more reliable predictive maintenance systems.

II. Background

A. Spectrogram

The spectrogram is a form of signal representation that provides a time-frequency visualization of the signal. Spectrograms are commonly used in literature works on anomaly detection, particularly in the context of audio signals. Their ability to highlight subtle changes in frequency content under varying operational conditions makes them a powerful tool for analyzing noisy industrial environments [7].

A spectrogram is computed using the Short-Time Fourier Transform (STFT), which analyzes the frequency content of a signal over short, overlapping time windows. The process involves segmenting the signal into frames, applying a window function to each frame to minimize spectral leakage, and computing the Fourier Transform on each windowed segment. In Figure 1, we present a spectrogram derived from a slider sound signal, which delineates the relationship between frequency and time.

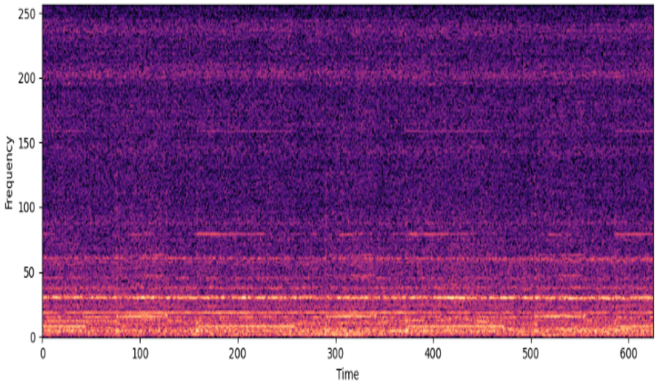

Fig. 1. Spectrogram (Slider) for 0 dB

The spectrogram *X(t, f)* calculation is defined in Equation (1) [7]:

$$X(t,f) = \sum_{n=0}^{N-1} x(n) \cdot w(n-t) \cdot e^{-j2\pi fn} \tag{1}$$

where:

- $x(n)$ is the input time-domain signal,
- $w(n-t)$ is the window function centered at time t,
- f is the frequency,
- and N is the window length.

The squared magnitude of the STFT coefficients results in a power spectrogram, which provides a visual representation of how the energy of different frequency components evolves. It makes the spectrogram especially valuable for detecting transient anomalies and identifying time-localized spectral features [3].

B. Mel Spectrogram

Recent studies have shown the effectiveness of Mel-scalograms combined with artificial neural networks for bearing fault diagnosis [10]. The Mel spectrum is an adequate representation for classifying various types of signals, including speech, music, and vibration signals (VS). Its extraction involves several key steps: signal framing, windowing, Fourier transform, application of the Mel filter bank, and logarithmic scaling. The signal is first divided into short, overlapping frames. A window function is applied to each frame to reduce spectral leakage. Then, the Fast Fourier Transform (FFT) is used to convert each windowed frame to the frequency domain. The resulting power spectrum is passed through a Mel filter bank, which maps the linear frequency axis to the perceptually motivated Mel scale [11].

The Mel spectrum $M(m)$ is computed as:

$$M(m) = \log\left(\sum_{k=1}^{K} |X(k)|^2 \cdot H_m(k)\right) \tag{2}$$

where:

- $X(k)$ is the magnitude of the Fourier-transformed signal at frequency bin k,
- $H_m(k)$ is the triangular filter corresponding to the m-th Mel filter,
- K is the number of frequency bins,
- and $\log$ denotes the natural logarithm.

The logarithmic operation enhances perceptually relevant components and compresses dynamic range. The Mel spectrum is conceptually similar to the scalogram produced by the Short-Time Fourier Transform (STFT) [11]. However, it differs in that it transforms the frequency axis into the Mel scale, which better reflects human auditory perception and emphasizes important patterns in low-frequency vibration signals.

This technique has been extensively used in signal classification tasks, including fault diagnosis in rotating machinery and audio processing domains.

C. MFCC Spectrogram

The Mel-Frequency Cepstral Coefficients (MFCC) technique is grounded in principles that emulate human auditory perception. It projects the spectral characteristics of a signal onto the Mel scale, a nonlinear frequency scale that is more

sensitive to lower frequencies. This makes MFCC particularly effective for extracting features from low-frequency components, where early signs of mechanical faults are often present. After mapping the power spectrum of each frame onto the Mel scale and applying a logarithmic transformation, a Discrete Cosine Transform (DCT) is performed on the resulting log-Mel spectrum to obtain the final Mel-frequency cepstral coefficients (MFCCs) [12].

In predictive maintenance applications, MFCC offers enhanced resolution in the low-frequency range, where acoustic emissions generated by early-stage defects, such as bearing faults, tend to concentrate. By amplifying weak and subtle signal components in these bands, the MFCC increases both the detectability and discriminative capacity of fault-related features. This capability is especially beneficial in scenarios involving low-speed and high-load conditions, such as in wind turbine pitch bearings, where fault signatures are typically low in energy. Consequently, MFCC plays a fundamental role in enabling the early identification of incipient failures, contributing to improved fault diagnosis and reduced downtime [12].

D. Kullback - Leibler Divergence

The Kullback-Leibler Divergence, also known as relative entropy [13], [14], is a statistical measure used to quantify the difference between two probability distributions. Typically, it is applied to compare an observed distribution with a reference or expected distribution. This divergence reflects the amount of information lost when the reference distribution is used to approximate the observed one and is widely employed in fields such as machine learning, signal processing, and predictive maintenance. For two continuous probability distributions P_1 and P_2, defined over a random variable x, the Kullback-Leibler Divergence is given by:

$$D_{KL}(P_1(x)\|P_2(x)) = \int_{-\infty}^{\infty} p_1(x) \log\left(\frac{p_1(x)}{p_2(x)}\right) dx \quad (3)$$

where $p_1(x)$ and $p_2(x)$ are the probability density functions associated with P_1 and P_2, respectively.

In the context of predictive maintenance, the Kullback-Leibler divergence has been used to identify deviations in signal behavior that may indicate early-stage faults in rotating machinery. For example, Wang *et al.* [15] demonstrated the application of this measure to detect anomalies by analyzing differences in probability density functions. Although their study focused on identifying human presence behind walls, the methodology is directly applicable to industrial environments, where subtle changes in vibration or acoustic signals can reveal the onset of mechanical degradation. Such approaches enhance the ability to perform early fault detection, reducing downtime and maintenance costs.

III. Materials and Methods

A. Experimental Setup and Data Processing

We used the Sound Dataset for Malfunctioning Industrial Machine Investigation and Inspection (MIMII dataset) in this research. This dataset presents both normal and anomalous sounds originating from various types of industrial machines, including sliders, valves, fans, pumps, and slide rails. Eight microphones collected the sound, recording 16-bit audio signals with a sample rate of 16 kHz [9].

We utilized this dataset to detect anomaly models based on machine learning. In this work, we focused on the slider records because they differentiate between normal and anomalous instances, as we observed in the Kullback Leibler divergence map presented in subsection IV-A.

Figure 2 illustrates the workflow of an audio-based anomaly detection system that applies to rotating machinery (slider) within the context of predictive maintenance.

The process starts with the acquisition of raw audio signals (waveforms), which we then transformed into three different feature representations: *Spectrogram*, *Mel Spectrogram*, and *Mel Frequency Cepstral Coefficients* (MFCCs). They extract these representations to compare their effectiveness in identifying anomalies. After extracting the representations, we created the Kullback-Leibler Divergence map to visually illustrate the differences in noise between anomalous and normal signals. This approach not only highlights their divergence but also helps us pinpoint which frequencies offer clearer distinctions between the two classes.

Next, the transformed data is fed into a *Training and Classification* module, where the system learns from the input data and performs binary classification. This process generates an output indicating whether the signal is normal or abnormal.

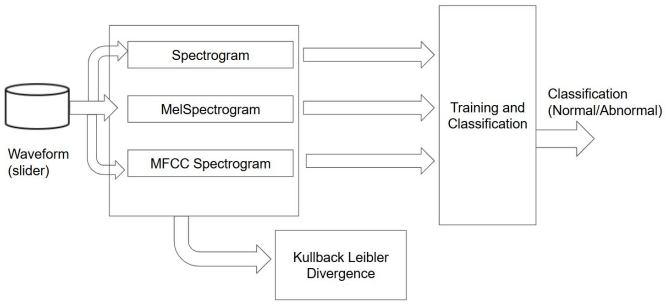

Fig. 2. Proposed Method

B. Experimental Design

To verify the best type of image for classifying anomaly and normal instances, we generated spectrograms, MelSpectrograms, and Mel-frequency spectral coefficients.

We generated spectrograms by applying the Short-Time Fourier Transform (STFT) to overlapping frames of the input signal using the Spectrogram function from the torchaudio.transforms library. We configured the transformation with the following parameters: n_fft = 1024, win_length = 1024, hop_length = 256, and power = 2. We consistently applied this configuration to all spectrograms used in our experiments.

Next, we extracted Mel spectrograms using the MelSpectrogram function from the torchaudio.transforms library. We set

the transformation parameters to sample_rate = 16000, n_fft = 1024, win_length = 1024, hop_length = 256, n_mels = 128, and power = 2.0. We applied this configuration uniformly to all input signals.

Additionally, we computed Mel Frequency Cepstral Coefficients (MFCCs) using the MFCC function. We configured it with sample_rate = 16000 and n_mfcc = 40, and we passed additional parameters through the melkwargs dictionary: n_fft = 1024, win_length = 1024, hop_length = 256, n_mels = 128, and power = 2.0. We consistently applied this setup to all MFCC representations used throughout the experiments.

We then inserted the images created from each method into a convolutional neural network (CNN). We generated and classified the spectrograms and then transformed them into arrays, which we used as input. We trained the Convolutional Neural Network (CNN) using the Cross-Entropy Loss function and optimized it with the Adam algorithm, setting the learning rate at 0.001. We conducted the training over 500 epochs, employing a batch size of 64 for training and a batch size of 1 for both validation and testing. The network architecture included a dense layer with 16 neurons. To construct the dataset, we used 25 normal and 25 anomalous spectrograms for training, while we reserved 25 normal and 25 anomalous samples for the test set. Subsequently, we allocated 20% of the test set for validation. The setup of the CNN included:

- **Input:** A single-channel (grayscale) image, typically a spectrogram or mel spectrogram with shape $1 \times H \times W$.
- **Convolutional Layer 1:**
  - `Conv2d` with 16 filters of size 3×3 and `padding="same"`
  - Activation function: ReLU
  - `MaxPool2d` with kernel size 2×2
- **Convolutional Layer 2:**
  - `Conv2d` with 32 filters of size 3×3 and `padding="same"`
  - Activation function: ReLU
  - `MaxPool2d` with kernel size 2×2
- **Convolutional Layer 3:**
  - `Conv2d` with 64 filters of size 3×3 and `padding="same"`
  - Activation function: ReLU
  - `MaxPool2d` with kernel size 2×2
- **Dropout:** Applied with a rate of 0.2 after convolutional layers to reduce overfitting.
- **Fully Connected Layer (FC):**
  - `Linear` layer with input size $64 \times 8 \times 8$ (flattened feature map) and output size defined by the parameter `dense_neurons` (e.g., 32).
- **Output Layer:**
  - `Linear` layer with 2 output neurons for binary classification.

We evaluated the performance of each model using the metrics of accuracy, precision, recall, and F1 score.

IV. Results

This section presents the results obtained in the experiments. First, we show the interclass maps using the Kullback-Leibler Divergence. Then, we present the results for the classification task.

A. Kullback–Leibler Divergence

The divergence maps in Figures 2, 3, and 4 display the Kullback-Leibler Divergence for 6dB, 0dB, and -6dB, respectively, showing that the interclass divergence becomes clearer at higher SNR levels. Specifically, when we work with a signal-to-noise ratio of 6 dB, the divergence becomes more pronounced. The three KL Divergence maps indicate that, for this element, the divergence is more expressive in the frequency range above 200 KHz.

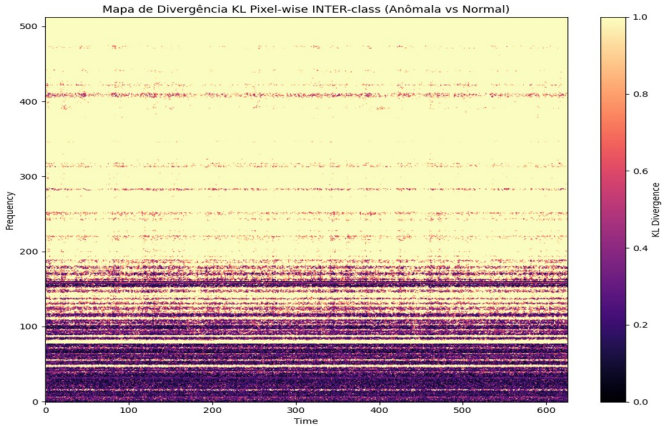

Fig. 3. Interclass Divergence Map (Anomalous vs Normal) for 6 dB

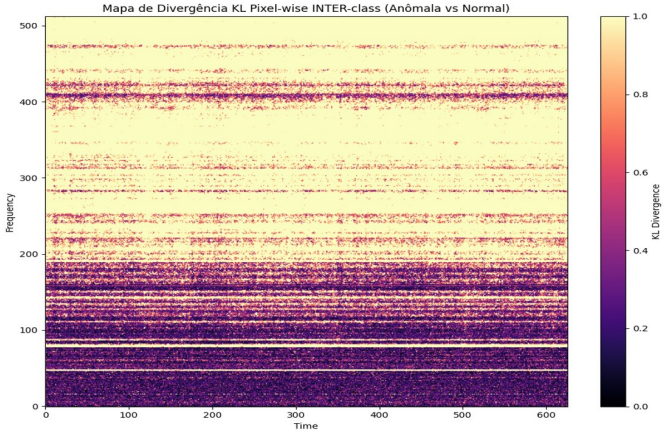

Fig. 4. Interclass Divergence Map (Anomalous vs Normal) for 0 dB

B. Classification

We inserted the spectrograms, including MelSpectrograms and MFCC Spectrograms, into the CNN to classify instances as anomalies or normal.

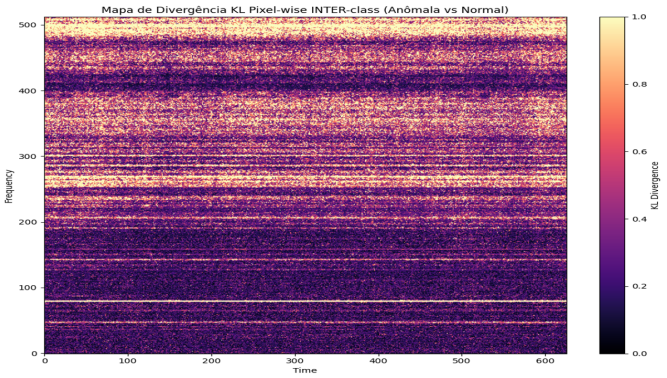

Fig. 5. Interclass Divergence Map (Anomalous vs Normal) for -6 dB

1) Performance Metrics: We classified each instance based on three different signal-to-noise ratios, as detailed in Tables I, II, and III. We conducted 30 experiments with varying initialization conditions, reporting the average values from these rounds, with the numbers in parentheses indicating the standard deviation. The Mel Spectrogram consistently produced the best results. The less variability of the standard deviation of the three SNRs demonstrates stability across all evaluation metrics, regardless of the signal-to-noise ratio (SNR) used. This fact demonstrates the model's excellent reliability.

These results highlight the key features of the Mel Spectrograms. The superior performance of Mel Spectrograms can be attributed to their perceptual scaling of frequency components, which emphasizes low-frequency regions where early mechanical faults typically manifest.

2) Statistical Significance and Robustness: The Wilcoxon test interclass and intraclass produced a p-value <0.05, which suggests that the data come from different distributions. This test demonstrates that we consistently achieved the best results for Mel Spectrogram, as detailed in IV. Table V shows significantly low p-values, also demonstrating that varying noise levels have a profound influence on the spectrograms.

In contrast, the MFCC Spectrograms displayed the poorest stability in the studied metrics, mainly because they are designed to process voice-based signals.

3) Accuracy Evolution During Training: Figures 6–8 illustrate the training and validation accuracy evolution over 500 epochs for each type of signal representation and SNR level. The Mel Spectrogram representation exhibits superior convergence speed and stability, with validation accuracy consistently above 85% even under severe noise conditions (-6 dB). In contrast, MFCC-based models exhibit marked instability and signs of overfitting, particularly as the signal-to-noise ratio (SNR) decreases. It suggests that MFCC, initially designed for speech processing, may fail to generalize under varying industrial acoustic profiles.

The Spectrogram-based models display intermediate behavior. They demonstrate reasonably high accuracy at 6 dB and 0 dB SNR levels, with more consistent validation performance than MFCCs. However, they are more susceptible to degradation under high-noise scenarios compared to Mel Spectrograms. The linear frequency scale and lack of perceptual compression make Spectrograms less effective in highlighting fault-relevant features, particularly in the presence of broadband noise.

These trends corroborate the findings summarized in Tables I–III and reinforce the suitability of Mel Spectrograms for robust and noise-resilient anomaly detection in industrial environments.

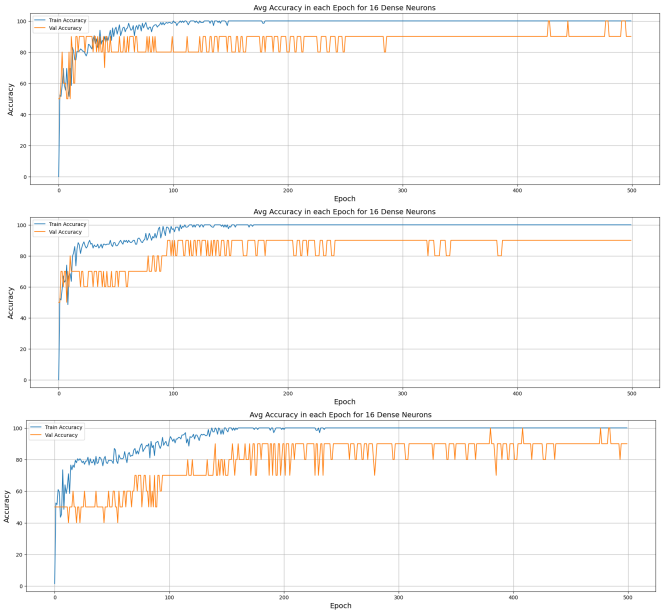

Fig. 6. Training and validation accuracy per epoch using Mel Spectrogram representation for SNR levels of 6dB (top), 0dB (middle), and -6dB (bottom). The Mel spectrogram exhibits high stability and rapid convergence across all conditions.

TABLE I
PERFORMANCE METRICS FOR SPECTROGRAM - SNR LEVELS

SNR	**6dB**	**0dB**	**-6dB**
ACC	0.98 (0,019)	0,88 (0,07)	0,85 (0,03)
F1 Score	0,98 (0,02)	0,86 (0,15)	0,85 (0,03)

TABLE II
PERFORMANCE METRICS FOR MFCC - SNR LEVELS

SNR	**6dB**	**0dB**	**-6dB**
ACC	0.94 (0.038)	0.85 (0.04)	0.74 (0.05)
F1 Score	0.94 (0.03)	0.84 (0.05)	0.72 (0.05)

TABLE III
PERFORMANCE METRICS FOR MELSPECTROGRAM - SNR LEVELS.

SNR	**6dB**	**0dB**	**-6dB**
ACC	0.99 (0,01)	0,93 (0,03)	0,87 (0,02)
F1 Score	0,99 (0,001)	0,93 (0,03)	0, 87 (0,02)

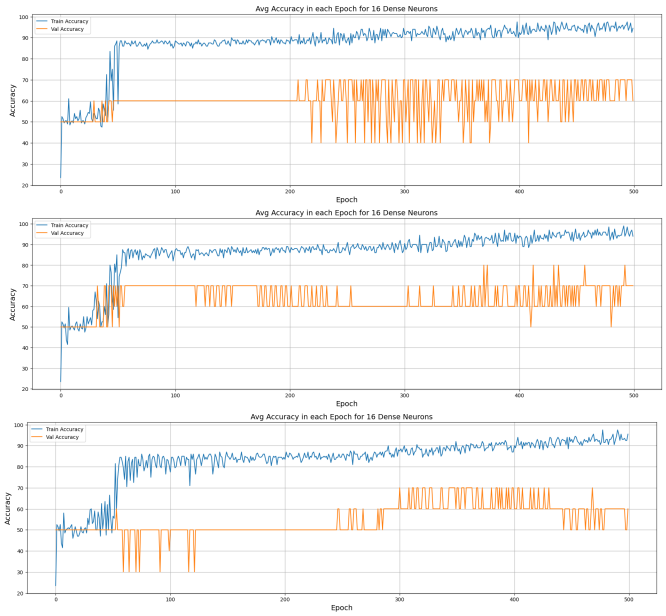

Fig. 7. Training and validation accuracy per epoch using MFCC representation for SNR levels of 6dB (top), 0dB (middle), and -6dB (bottom). Despite achieving a decent performance at 6 dB, MFCC is more sensitive to noise and exhibits unstable validation accuracy at lower SNRs.

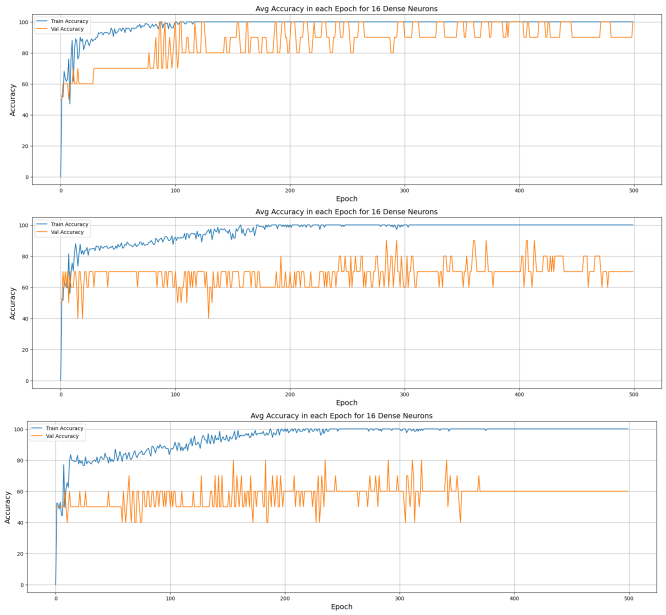

Fig. 8. Training and validation accuracy per epoch using standard Spectrogram representation across different SNR levels. Results indicate moderate robustness, but with higher variance than Mel Spectrogram.

TABLE IV
WILCOXON TEST OF METHODS ACROSS DIFFERENT NOISE LEVELS

Comparison	Noise Level	Statistic	P-value
MFCC vs Spectrogram	6dB	6.5	0.000087
MFCC vs MelSpectrogram	6dB	0.0	0.000003
Spectrogram vs MelSpectrogram	6dB	11.0	0.003816
MFCC vs Spectrogram	0dB	58.5	0.004906
MFCC vs MelSpectrogram	0dB	0.0	0.000001
Spectrogram vs MelSpectrogram	0dB	41.0	0.001046
MFCC vs Spectrogram	-6dB	0.0	0.000002
MFCC vs MelSpectrogram	-6dB	4.0	0.000005
Spectrogram vs MelSpectrogram	-6dB	40.5	0.02

TABLE V
WILCOXON TEST RESULTS COMPARING DIFFERENT NOISE LEVELS ACROSS REPRESENTATION TYPES.

Spectrogram Type	Comparison	Statistic	P-value
MFCC	6dB vs 0dB	0.0	0.000003
MFCC	6dB vs -6dB	0.0	0.000002
MFCC	0dB vs -6dB	0.0	0.000002
Spectrogram	6dB vs 0dB	0.0	0.000002
Spectrogram	6dB vs -6dB	0.0	0.000001
Spectrogram	0dB vs -6dB	65.5	0.001560
MelSpectrogram	6dB vs 0dB	0.0	0.000003
MelSpectrogram	6dB vs -6dB	0.0	0.000002
MelSpectrogram	0dB vs -6dB	11.0	0.00001

V. CONCLUSIONS

We aimed to determine the most effective representation of acoustic signals for studying anomaly detection in industrial machines using the results of the transformations for feeding a deep-learning classifier. After analyzing metrics such as accuracy, precision, F1-score, and recall, we found that the Mel Spectrogram serves as the most effective representation. We observed that values such as accuracy and recall at an SNR of 6 dB were 0.99 and 0.98, respectively, highlighting the performance of the MelSpectrogram. With the lower SNR, we were also able to note that the Mel Spectrogram achieved strong results in terms of accuracy and recall, with 0.87 and 0.90, demonstrating the consistency in the metrics for that representation.

We considered this step vital for paving the way to construct models that enhance predictive maintenance in industrial environments, especially those operating in noisy environments. Based on our findings, we will explore a new methodological process centered on the Mel Spectrogram. Our current work employs a single component, the slider, but we plan to analyze additional datasets in future research to test the model's generalization. We will also consider other classification methods, along with the exploration of various types of representations, such as scalograms. Additionally, we will examine data augmentation.

ACKNOWLEDGMENT

The authors would like to express their gratitude for the support received during the development of this work, which originated from the international collaboration promoted by the project STIC AmSud Nro. 23-STIC-06. In addition to

thanking the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001 for the financial support, which enabled the advancement of this research.

REFERENCES

[1] SERRADILLA, Oscar; ZUGASTI, Ekhi; RODRIGUEZ, Jon; ZURUTUZA, 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.

[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; RENÒ, Vito. A review of deep learning based anomaly detection strategies in Industry 4.0 focused on application fields, sensing equipment and algorithms. *IEEE Access*, 2024.

[5] DE OLIVEIRA NETO, Wilson A.; GUEDES, Elloá B.; FIGUEIREDO, Carlos Maurício S. Anomaly Detection in Sound Activity with Generative Adversarial Network Models. *Journal of Internet Services and Applications*, v. 15, n. 1, p. 313–324, 2024.

[6] MERANEH, Awaleh Houssein; AUTREL, Fabien; BOUDER, Hélène Le; PAHL, Marc-Oliver. SADIS: real-time sound-based anomaly detection 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, Ruixiang; YAO, Fangzhi; XIONG, Junhan; QIU, Hongbing. A machine anomalous sound detection method using the LMS spectrogram and ES-MobileNetV3 network. *Applied Sciences*, v. 13, n. 23, p. 12912, 2023.

[9] 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

[10] S. A. Taye, F. S. Alqahtani, and A. M. Alanazi, *Advanced bearing-fault diagnosis and classification using Mel-scalograms and FOX-optimized ANN*, Sensors, vol. 23, no. 3, p. 1444, 2023.

[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.

[13] S. Kullback and R. A. Leibler, "On Information and Sufficiency," *Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, 1951.

[14] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed., Wiley-Interscience, 2006.

[15] H. Wang, et al., "Anomaly Detection using KL Divergence in Signal Processing," *Journal of Applied Signal Processing*, 2020.