forum acusticum 2026
8-12 September · Graz, Austria

INTERPRETABLE ANOMALY DETECTION IN ROTATING MACHINES FOR EXPERTS USING CONVOLUTIONAL AUTOENCODER

Anouck BRUGUIERE[1,2], Kais HASSAN[1], Jean-Hugh THOMAS[1], Amaury GRAILLAT[2]

[1] *Laboratoire d'Acoustique de l'Université du Mans (LAUM), UMR 6613, Institut d'Acoustique-Graduate School (IA-GS),CNRS, Le Mans Université, 72085 Le Mans, France*
[2] *OROS Digital, 82 Allee Galilee, 38330 Montbonnot-Saint-Martin, France*

This paper is a revised version of the authors' submitted manuscript that was peer-reviewed by at least two anonymous reviewers.

Abstract

Unplanned breakdowns in industrial rotating machinery can cost millions per incident, making fault detection a critical maintenance strategy. This work presents an anomaly detection approach based on a 2D convolutional autoencoder (CAE) trained exclusively on healthy vibration data. The publicly available Mechanical Faults in Rotating Machinery (MFRM) dataset is used, comprising acceleration signals acquired under normal conditions and three fault types: unbalance, misalignment, and mechanical looseness. Vibration signals are first resampled into the angular domain via Computed Order Tracking (COT), then transformed into order spectrograms (OS) whose axes, harmonic orders and shaft revolutions, are directly meaningful to domain experts. A Log-Z standardisation enhances the sensitivity of the CAE to subtle spectral deviations. Trained to minimise reconstruction error on healthy OS only, the OS-CAE produces elevated mean squared error (MSE) when encountering faulty inputs. A threshold calibrated on the combined training and validation MSE distributions separates healthy from anomalous samples. Beyond binary detection, a per-order MSE profile of reconstructed spectrograms localises reconstruction failures at specific harmonic orders, providing interpretable visual cues that hint at the nature of the detected fault without relying on post-hoc explainability methods.

Keywords: *vibration analysis, rotating machines, anomaly detection, convolutional autoencoder, order spectrogram*

***Corresponding author**: anouck.bruguiere@oros.com.*

1. Introduction

1.1. Context

Industrial rotating machinery, such as turbines (water, gas), compressors, and pumps, are critical pieces of equipment whose failure can lead to production stoppages, material damage, and, more rarely, serious accidents with dramatic human and economic consequences.[1] One of the most used solutions in maintenance and risk prevention is vibration monitoring. Empirical methods and norms (ISO-10816) rely on signal processing and statistical techniques that compare vibration measurements to predefined thresholds determined by analyzing the characteristic parameters of healthy and defective states of the rotating machines. However, current monitoring systems generate enormous volumes of complex, predominantly healthy, causing class imbalance, and unlabeled data. That reality maintains a dependence on human expertise, which is difficult to mobilize and costly, and which is itself overwhelmed by the flood of data.

Existing artificial intelligence methods, including machine learning (ML) and deep learning (DL), can assist experts and help refine their diagnoses. However, they have two major limitations. On the one hand, traditional algorithms such as Support Vector Machine (SVM) depend entirely on the quality of the features extracted by an expert, a long term process that also requires in-depth knowledge of system mechanics and signal processing. This dependency restricts the adaptability of models [1][2]. On the other hand, DL can automatically extract features for end-to-end diagnosis, but requires extensive labeled datasets that are rarely available in real-world industrial settings and lacks explainability [3]. Additionally, DL models function as "black boxes," making their decisions not easily

[1] For example, the explosion of the turbine at the Saiano-Shushenskaya hydroelectric power plant in 2009, caused by undetected vibrations, resulted in the deaths of 75 people and damage estimated at several hundred million dollars.

explainable. This hinders their adoption in industrial environments where trust and interpretability are crucial [4]. The challenge is therefore to design systems that are not only powerful and robust, but also reliable and interpretable.

1.2. Anomaly detection

1.2.1. Rare event detection

Machinery fault detection is classified within the R1 rarity level, representing extreme rarity with a fault event frequency typically between 0% and 1% [5]. Predictive maintenance and high surveillance ensure that faults in industrial rotating machinery remain rare; ironically, this very success makes them even more difficult to identify. This phenomenon is known as the Curse of Rarity (CoR) [6]. Consequently, because safety-critical events are so infrequent, standard deep learning models often fail to find enough fault signal to learn from. The technical challenge lies in how models update during training. At each step, a model estimates the direction in which it should update its parameters (the gradient $\tilde{\mu}$) based on a stochastic batch of data. $\tilde{\mu}$ is defined as:

$$\tilde{\mu} \triangleq \frac{1}{n} \sum_{i=1}^{n} \nabla_\theta \, f_\theta(X_i) \tag{1}$$

where θ is the parameters f_θ is the loss function, X_i a sample from the training distribution, and n the batch size. In our case, the vast majority of X_i samples belong to a set A of "normal condition" or "healthy" events. These normal events lack information about potential faults, hence function as noise that drowns out the rare "fault" signals, causing the variance of the gradient estimation to explode. This issue can be solved by reversing the approach. Instead of training a model on rare fault signature, we focus exclusively on what is known the "healthy" set A. By mastering the representation of set A, the model becomes highly specialized, and no longer relies on the scarce, high-variance samples in faulty set B to define the model's primary parameters.

1.2.2. Autoencoders

This study focuses on semi-supervised anomaly detection models, in which an autoencoder is trained on normal data and anomalies are detected based on the reconstruction error. In the context of rotating machinery fault diagnosis, Principi *et al.* [7] applied unsupervised deep autoencoders to electric motor vibration signals. Their work showed that autoencoders trained on healthy motor data can detect incipient faults, ideal for industrial condition monitoring.

1.2.3. Convolutional autoencoder

Order spectrograms are two-dimensional (2D) inputs, hence a Convolutional neural networks (CNN) are a natural choice: their learnable filters capture local spectral patterns while preserving the 2D structure of the spectrogram. A convolutional autoencoder (CAE) combines this spatial feature extraction with dimensionality reduction, have shown effective feature learning from image input [8] and have been applied in rotating machines by Hasan *et al.* [9]. They demonstrated that spectral representations capture fault signatures more effectively than raw time-domain features, particularly when operating conditions vary. Furthermore, a key choice in this work is the use of OS as input. COT resamples the signal in the angular domain, so that harmonics of the shaft speed appear at fixed *orders*. Wang *et al.* [10] combined COT with deep learning, and showed that order-domain representations improve fault classification accuracy compared to time-frequency approaches. Although the MFRM dataset operates at near-constant speed, the shaft speed was maintained at approximately 1238 rpm, the order representation remains advantageous here for a different reason: each row of the spectrogram corresponds to a specific harmonic order, making reconstruction errors directly interpretable in terms of known fault signatures (e.g., an abnormally high amplitude at the first order (1X) amplitude indicates unbalance, see Sections 2.1 and 2.2.4).

2. Methodology

2.1. MFRM

The MFRM dataset [11], introduced by Brito *et al.*, consists of acceleration signals acquired from a test bench under normal conditions and three mechanically induced fault types. Unbalance refers to an uneven mass distribution on the rotor, generating a centrifugal force synchronous with the shaft speed. Misalignment occurs when the shaft axis deviates from its intended geometric position relative to the bearings or coupling, in the angular or parallel direction. Mechanical looseness is caused by insufficient fastening of structural components, allowing excessive play. Vibration signals are sampled at 25.6 kHz with 25 000 points per acquisition from four accelerometers. Only accelerometer 1 is used, following [12]. The dataset contains 20 test sessions (5 per class), each comprising 420 signals per accelerometer acquired continuously (∼6.8 min at ∼1 s per signal). Before each session, the bench was disassembled and reassembled, introducing subtle mechanical differences (belt tension, bearing seating, assembly tolerances).

On the same dataset, Brito *et al.* [12] proposed an unsupervised anomaly detection methodology, combining time and frequency domain feature extraction with several ML algorithms (Isolation Forest, One-Class SVM etc.), achieving F1-scores up to 99.22% and area under the precision-recall curve (PR-AUC) up to 99.96%. For explainability, they employed SHapley Additive exPlanations (SHAP) to rank feature importance. In contrast, we pursue interpretability by relying on order spectrograms as input features, a representation that domain experts are already familiar with.

2.2. Signal pre-processing

2.2.1. Data split protocol

The data split follows a session-based separation between training, validation and evaluation (Table 1). Healthy test data is drawn from sessions 03 and 06, which were recorded on different days and thus exhibit distinct structural properties due to reassembly (belt tension, bearing seating). Including these sessions in the test set evaluates the robustness of the CAE to inter-session variability and verifies that normal mechanical differences do not trigger false alarms.

Table 1: Data split protocol based on session and class. Signal counts refer to the number of acquisitions from accelerometer n°1 per session.

Role	Sessions	Class	Signals
Training	01, 09	Healthy	840
Validation	17	Healthy	420
Test	03, 06	Healthy	420
Test	04	Unbalance	140
Test	02	Misalignment	140
Test	05	Looseness	140

2.2.2. Angular resampling (COT)

The first step converts the signal from the time domain to the angular domain using COT [10]. The instantaneous angular position of the shaft for constant speed is computed as:

$$\varphi(t) = 2\pi f_r \, t \tag{2}$$

A uniformly spaced angular grid φ_k is then constructed with a resolution of N_{spr} samples per revolution (set to 1024 in this work). The original signal $x(t)$ is interpolated onto this grid to produce the angularly resampled signal $x_a(\varphi)$. As a result, one shaft revolution always spans exactly N_{spr} samples.

2.2.3. Order spectrogram via STFT

A Short-Time Fourier Transform (STFT) is applied to $x_a(\varphi)$ using a Hann window $w[n]$ of length $N_w = 2 \cdot N_{\text{spr}} = 2048$ samples (two revolutions) with 75% overlap (hop size $H = N_w/4$). The order spectrogram is defined as the squared magnitude of the STFT:

$$S(m,k) = \left| \sum_{n=0}^{N_w-1} x_a[m \cdot H + n] \cdot w[n] \cdot e^{-j2\pi kn/N_w} \right|^2 \tag{3}$$

where m is the angular window index and k the order bin. Since the sampling rate of x_a is N_{spr} samples per revolution, k is naturally expressed in *orders* (multiples of f_r) rather than Hertz, with a resolution of $\Delta k = 0.5$ order. At 1238 rpm, ($f_r \approx 20.63$ Hz), each 1-second acquisition yields ∼20.6 revolutions, producing a spectrogram of $H = 17$ order bins (orders 0 to 8) and $W = 43$ angular frames. Each signal is thus represented as a 17×43 matrix constituting the input $\mathbf{x} \in \mathbb{R}^{H \times W}$ to the OS-CAE. The limited order resolution means that fault signatures may appear slightly offset from their theoretical integer harmonic.

2.2.4. Standardization

Log-scaled spectrograms compress the dynamic range and allow autoencoders to learn subtle spectral deviations rather than focusing on dominant peaks, as demonstrated for industrial anomalous sound detection [13]. In the present work, a combined Log-Z standardisation is applied to the order spectrograms. First, a logarithmic transformation $\log(1 + |X|)$ compresses the dynamic range. Any log-amplitude exceeding the 99.5th percentile of the training distribution (i.e., the value below which 99.5% of training pixels fall) is capped to that value, attenuating extreme outliers:

$$X_{\text{clip}} = \min\Big(\log(1+|X|),\ Q_p\big(\log(1+|X_{\text{train}}|) \big) \Big) \tag{4}$$

Then, Z-score standardisation is applied using the mean $\mu_{\log}$ and standard deviation $\sigma_{\log}$ of the clipped log-amplitudes, computed exclusively on the training set (sessions 01 and 09, see Table 1):

$$X_{\text{norm}} = \frac{X_{\text{clip}} - \mu_{\log}}{\sigma_{\log}} \tag{5}$$

A small constant (10^{-8}) is added to $\sigma_{\log}$ for numerical stability. This Log-Z strategy was selected over min–max, Z-score and no standardisation after hyperparameter search. As will be discussed in Section 3.2, Figure 2a–d presents order spectrograms for each machinery condition (Row 1). Each class exhibits a distinct spectral signature: unbalance is dominated by 1X (Figure 2a), misalignment by 2X (Figure 2b), and looseness by both 1X and 3X (Figure 2c).

2.3. Architecture

2.3.1. Autoencoder

The autoencoder (AE) is a neural network composed of two symmetrical parts: an encoder e compresses the input $\mathbf{x}$ into a low-dimensional latent representation $\mathbf{z} = e(\mathbf{x})$, and a decoder d reconstructs an approximation $\mathbf{x}' = d(\mathbf{z})$ from this compressed representation [14]. Each input $\mathbf{x} \in \mathbb{R}^{H \times W}$ is an angular order spectrogram, where H denotes the number of order bins (0 to N_{ord}) and W the number of angular frames per acquisition. The network is trained by minimising the mean squared error (MSE) between input and reconstruction over the training set:

$$\mathcal{L}_{\text{AE}} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{H \cdot W} \left\| \mathbf{x}_i - \mathbf{x'}_i \right\|^2 \tag{6}$$

where N is the number of training samples, $\mathbf{x'}_i = d(e(\mathbf{x}_i))$, and $\| \cdot \|^2$ denotes the squared Euclidean norm over the $H \times W$ spectrogram pixels.

2.3.2. Convolutional autoencoder

The model used in this work is a CAE that replaces the dense layers of a standard AE with convolutional and transposed-convolutional layers, following the architecture proposed by Masci *et al.* [8], and operates on OS as illustrated in Figure 1. Reconstruction is

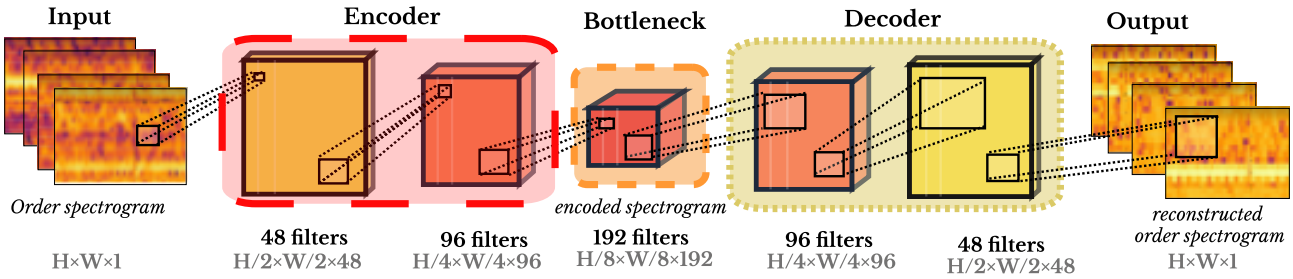

Figure 1: Architecture of the convolutional autoencoder. The encoder compresses the input order spectrogram through successive convolutional layers (*stride 2, batch normalisation, ReLU*) into a low-dimensional bottleneck, and the decoder reconstructs the original input via transposed convolutions. The final layer has no activation function. The Latent space dimensions are:$H/8 \times W/8 \times 192$.

cropped to the input dimensions to handle non-integer spatial sizes arising from stride-2 convolutions.

2.3.3. Detection thresholds

The OS-CAE is trained to minimise the MSE (Equation (6)) between the input spectrogram and its reconstruction. At inference time, the per-sample MSE between input and reconstruction serves as the anomaly score:

$$\text{score}(\mathbf{x}) = \frac{1}{H \cdot W} \left\| \mathbf{x} - \mathbf{x}' \right\|^2 \tag{7}$$

A healthy signal yields a low score, while a faulty signal produces a high score, as the OS-CAE cannot reconstruct patterns it has never seen. To account for inter-session variability due to reassembly differences, the detection threshold is calibrated on the combined MSE distribution of the training and validation sets:

$$\tau = P_q \left(\{\text{score}(\mathbf{x}_i)\}_{i \in \text{train} \cup \text{val}} \right) \tag{8}$$

where P_q denotes the q-th percentile. With $q = 95$, approximately 5% of healthy samples are expected to exceed the threshold.

2.4. Evaluation & Metrics

2.4.1. Scores

Performance is evaluated on unseen test data using standard classification metrics derived from the confusion matrix: true positives (TP, correctly detected faults), true negatives (TN, correctly identified healthy samples), false positives (FP, false alarms), and false negatives (FN, missed faults). Accuracy (Eq. 9) measures the ratio of correct predictions over all samples, but is deceptive under extreme class imbalance: in an R1 rarity scenario, a trivial "always healthy" predictor would exceed 99% accuracy while missing every fault.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

Precision and Recall address this limitation by focusing on the minority class. Precision quantifies the reliability of alarms, while Recall measures the capacity to avoid missed faults:

$$\text{Precision} = \frac{TP}{TP + FP} \quad ; \quad \text{Recall} = \frac{TP}{TP + FN} \tag{10}$$

The F1-Score balances both through their harmonic mean:

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{11}$$

2.4.2. MSE per-order profile

To provide interpretability, the per-sample MSE is decomposed along the order axis by averaging the squared reconstruction error over all angular frames for each order bin. These maps localise which harmonic orders trigger the anomaly alerts, offering domain experts a direct visual cue for fault diagnosis.

3. Results

3.1. Implementation details

The OS-CAE is optimised using Adam with an initial learning rate of 5×10^{-4}, L_2 weight decay of 10^{-4}, and a batch size of 16 over a maximum of 1 200 epochs. Early stopping with a patience of 200 epochs prevents overfitting, and the model state corresponding to the lowest validation loss is retained (epoch 427, val. loss $= 4.56 \times 10^{-3}$). Input spectra are normalised via a log-Z scheme ($\mu = 1.69 \times 10^{-3}$, $\sigma = 1.21 \times 10^{-3}$, clip $= 7.16 \times 10^{-3}$). These hyper-parameters (number of filters, learning rate, weight decay, batch size, number of orders) were selected by random search [15]. All experiments were conducted on a laptop equipped with an AMD Ryzen 9 6900HS processor (8 cores, 3.30 GHz base clock) and 32 GB of RAM, running a 64-bit operating system.

3.2. Detection performance

Following standard industrial practice, two threshold levels are defined from eq.(7) (Table 2).

Table 2: Detection performance of the OS-CAE under two industrial thresholding levels.
T = Threshold, FP = False alarms, FN = Missed faults. Test set: 420 healthy + 420 faulty order spectrograms.

Level	T	FP	FN	F1 (%)
Alert (P95)	0.00585	5	0	99.41
Alarm (μ+2σ)	0.00647	1	0	99.98

The *alert* level (P95, $\tau = 0.00585$) provides early warning with an F1-score of 99.41% and perfect Recall,

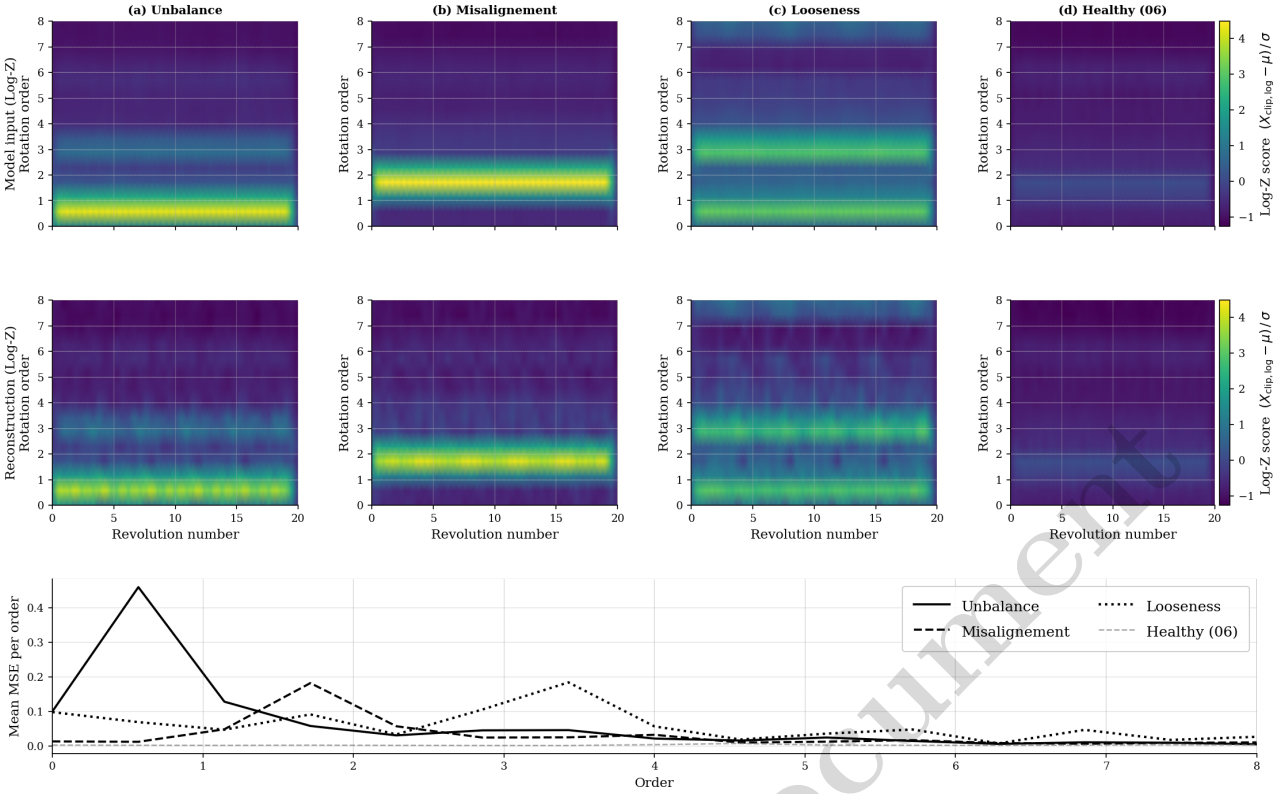

Figure 2: Order spectrogram reconstruction by the OS-CAE for each fault type and the healthy reference (Test06). *Row 1*: input $\mathbf{x}$. *Row 2*: reconstruction $\mathbf{x}'$. *Row 3*: mean pixel-wise MSE per order, averaged over all angular frames. Unbalance peaks at 1X, misalignment at 2X, looseness shows a broadband profile around 3X–4X. The healthy reference (dashed grey) remains near zero, confirming great reconstruction of normal patterns.

but 5 false alarms on Test03. Frequent false alarms can lead to alert fatigue, where technicians gradually lose trust in the system and may start ignoring warnings. The stricter *alarm* level (μ+2σ, $\tau = 0.00647$) mitigates this risk by reducing false alarms to 1 while maintaining Recall at 100% and raising F1 to 99.98%. In both cases, no fault is missed (FN = 0), confirming that the OS-CAE cleanly separates the healthy and faulty distributions (separability = 68σ) despite inter-session variability.

3.3. Fault-type analysis

Table 3: Reconstruction error (MSE $\times 10^{-3}$) statistics across operational states. *Session numbers in parentheses. Peak order (PO): order with highest reconstruction error.*

Set	Class	Mean	Std	PO
Training	Healthy (01)	5.27	0.61	–
	Healthy (09)	3.13	0.60	–
Validation	Healthy (17)	4.55	0.60	–
Test	Healthy (03)	4.78	0.50	–
	Healthy (06)	3.17	0.50	–
Test	Unbalance	**80.33**	4.97	1X
	Misalignment	**43.39**	4.18	2X
	Looseness	**82.35**	8.51	3X

Table 3 reports reconstruction error statistics across operational states. The mean MSE of any unseen healthy session remains below 4.78×10^{-3} (Test03), while all three fault types largely exceed this bound with zero overlap (separability = 68σ).

3.3.1. Unbalance

It produces the highest mean error (80.33×10^{-3}), 17× the healthy bound, driven by a massive energy concentration near the 1X order.

3.3.2. Looseness

It reaches a comparable level (82.35×10^{-3}) but with a broadly distributed profile spanning integer harmonics (1X, 3X) and fractional sub-harmonics, consistent with the excitation of the whole structure.

3.3.3. Misalignment

It is the subtlest fault (43.39×10^{-3}), yet still 9× above the healthy bound. The reconstruction error concentrates around the 2X harmonic order, characteristic of angular misalignment.

All false alarms originate from Test03 healthy signals. Misalignment is the lowest MSE fault. It suggests that milder cases could fall below the detection threshold. Testing with additional misalignment sessions would be valuable.

3.4. Visual interpretability

While Table 3 quantifies the global separation between healthy and faulty distributions. Figure 2 presents the input order spectrograms $\mathbf{x}$ (Row 1), OS-CAE reconstructions $\mathbf{x}'$ (Row 2), and the profile mean MSE

per order (Row 3) for each fault type and one healthy reference (Test06). The healthy reference (dashed grey in Row 3) remains near zero across all orders, confirming faithful reconstruction of normal patterns.

3.4.1. Unbalance

Figure 2a exhibits a peak near 1X, consistent with the synchronous force generated by mass imbalance.

3.4.2. Looseness

Figure 2c shows a broadly distributed MSE profile across multiple harmonics and sub-harmonics, reflecting the excitation of the whole structure.

3.4.3. Misalignment

Figure 2b concentrates its error around 2X, characteristic of angular misalignment.

The discriminant orders identified by the per-order MSE (1X for unbalance, 2X for misalignment, broadband for looseness) match the theoretical spectral signatures of these faults, confirming that OS-CAE reconstruction errors reflect the underlying physics and are directly interpretable without post-hoc explainability methods.

4. Conclusion

This work demonstrates that a convolutional autoencoder (OS-CAE) trained exclusively on healthy order spectrograms shows robust anomaly detection on the MFRM dataset (F1 = 99.41% at alert level, 99.98% at alarm level, with zero missed faults) while offering interpretability accessible to experts. In fact, harmonics orders make it possible to quickly identify the type of fault based on their characteristic signature: 1X for imbalance, 2X for misalignment, and a broadband signal for looseness. It should be noted, however, that interpretability does not equate to explainability. While the order-spectrogram representation makes the *outputs* of the model transparent to domain experts, the internal reconstruction process of the OS-CAE remains a black box. Full trustworthiness would necessitate complementing this approach with explainability (XAI) methods that expose the model's internal decision process.

From an industrial standpoint, this approach requires only healthy data for training, which is the most readily available condition in monitored installations. The anomaly score (per-sample MSE) is a single, intuitive indicator compatible with standard alert/alarm thresholding practices. Moreover, the per-sample MSE can be trended over successive acquisitions as a health indicator. De Fabritiis and Gryllias [16] showed that such scores progressively increase with degradation, bridging anomaly detection and predictive maintenance. Future work will extend this study to variable-speed operating conditions, incipient fault severities, multi-sensor fusion, and validation on industrial field data.

References

[1] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, and A. K. Nandi, "Applications of machine learning to machine fault diagnosis: A review and roadmap," *Mech. Syst. Signal Process.*, vol. 138, p. 106 587, 2020.

[2] H. Li, Z. Zhang, T. Li, and X. Si, "A review on physics-informed data-driven remaining useful life prediction," *Mech. Syst. Signal Process.*, vol. 209, p. 111 120, 2024.

[3] W. Cui, G. Meng, A. Wang, X. Zhang, and J. Ding, "Application of rotating machinery fault diagnosis based on deep learning," *Shock Vib.*, vol. 2021, p. 3 083 910, 2021.

[4] O. Das, D. B. Das, and D. Birant, "Machine learning for fault analysis in rotating machinery: A comprehensive review," *Heliyon*, vol. 9, no. 6, 2023.

[5] C. Shyalika, R. Wickramarachchi, and A. Sheth, "A comprehensive survey on rare event prediction," *ACM Comput. Surv.*, 2024.

[6] H. X. Liu and S. Feng, "Curse of rarity for autonomous vehicles," *Nat. Commun.*, vol. 15, p. 4808, 2024.

[7] E. Principi, D. Rossetti, S. Squartini, and F. Piazza, "Unsupervised electric motor fault detection by using deep autoencoders," *IEEE/CAA J. Autom. Sin.*, vol. 6, no. 2, pp. 441–451, 2019.

[8] J. Masci, U. Meier, D. Ciresan, and J. Schmidhuber, "Stacked convolutional auto-encoders for hierarchical feature extraction," in *Proc. ICANN*, 2011, pp. 52–59.

[9] M. J. Hasan, M. M. M. Islam, and J.-M. Kim, "Acoustic spectral imaging and transfer learning for bearing fault diagnosis," *Measurement*, vol. 175, p. 109 077, 2021.

[10] Y. Wang, Z. Zhang, X. Ding, Y. Du, J. Li, and P. Chen, "A reference learning network for fault diagnosis of rotating machinery under strong noise," *Applied Soft Computing.*, 2024.

[11] L. Brito, G. A. Susto, J. N. Brito, and M. Duarte, *Mechanical faults in rotating machinery dataset*, Mendeley Data, V3, 2023.

[12] L. C. Brito, G. A. Susto, J. N. Brito, and M. A. V. Duarte, "An explainable AI approach for monitoring mechanical faults in rotating machinery," *Mech. Syst. Signal Process.*, vol. 181, p. 109 462, 2022.

[13] Y. Kawaguchi et al., "Description and discussion on DCASE 2021 challenge task 2," in *Proc. DCASE Workshop*, 2021.

[14] U. Michelucci, *An introduction to autoencoders*, 2022. arXiv: 2201.03898.

[15] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *J. Mach. Learn. Res.*, vol. 13, pp. 281–305, 2012.

[16] F. De Fabritiis and K. Gryllias, "A self-supervised learning approach for anomaly detection in rotating machinery," in *Proc. PHM Society European Conf.*, 2024.