Contents lists available at ScienceDirect

Expert Systems With Applications

Fault Diagnosis using eXplainable AI: A transfer learning-based approach for rotating machinery exploiting augmented synthetic data

Lucas Costa Brito [a],*, Gian Antonio Susto [b], Jorge Nei Brito [c], Marcus Antonio Viana Duarte [a]

[a] *School of Mechanical Engineering, Federal University of Uberlândia, Av. João N. Ávila, 2121, Uberlândia, Brazil*
[b] *Department of Information Engineering, University of Padova, Via Gradenigo 6/B, 35131, Padova, Italy*
[c] *Department of Mechanical Engineering, Federal University of São João del Rei, P.Orlando, 170, São João del Rei, Brazil*

ARTICLE INFO

ABSTRACT

Due to the growing interest for increasing productivity and cost reduction in industrial environment, new techniques for monitoring rotating machinery are emerging. Artificial Intelligence (AI) is one of the approaches that has been proposed to analyze the collected data (e.g., vibration signals) providing a diagnosis of the asset's operating condition. Three major problems hinder the application of AI models in industrial environments, being the motivation for the research: (i) impossibility or long time to obtain a sample of all operational conditions (since faults rarely happen), (ii) high cost of experts to label all acquired data (in a supervised scenario) and (iii) lack of interpretability of the models (black-boxes). To overcome these problems, a new generic and interpretable approach for classifying faults in rotating machinery based on transfer learning from augmented synthetic data to real rotating machinery is here proposed, namely FaultD-XAI (Fault Diagnosis using eXplainable AI). To provide scalability using transfer learning, synthetic vibration signals are created mimicking the characteristic behavior of failures in operation. The application of Gradient-weighted Class Activation Mapping (Grad-CAM) with 1D Convolutional Neural Network (1D CNN) allows the post hoc interpretation of results, supporting the user in decision making and increasing diagnostic confidence. The proposed approach not only obtained promising diagnostic performance but was also able to learn characteristics used by experts to identify conditions in a source domain and apply them in another target domain. The experimental results obtained on three datasets containing different mechanical faults suggest the method offers a promising approach on exploiting transfer learning, synthetic data and explainable artificial intelligence for fault diagnosis. Lastly, to guarantee reproducibility and foster research in the field, the developed dataset is made publicly available.

1. Introduction

Currently, with the great need to increase the amount of final product manufactured, the industry has been looking for ways to monitor its assets in order to avoid unexpected breaks that can directly impact production (Brito et al., 2022). Due to the growing demand, searches for alternatives in the monitoring of rotating machinery have been commonplace, leading to a large amount of information and research being generated: from the definition of informative signals to the development of smart data processing techniques, from new sensors to new best practice in data acquisition and monitoring (Asutkar et al., 2023).

The increase in information generates a need for experts to achieve quick analysis and effective diagnosis. Thus, artificial intelligence models have been proposed to assist in the diagnosis and monitoring of assets.

For the task of monitoring rotating machinery, AI models have high accuracy rates when trained in supervised condition, ie. in the presence of labeled data (Lei et al., 2020) representing both normal and faulty conditions. Unfortunately, the supervised scenario is difficult to be applied in real world industrial settings, for two main reasons (Yang et al., 2019):

(i) *Faults rarely happen in real world*: by design, equipment is made to be ideally operating all the time in a normal condition, therefore not providing sample labels for faulty conditions. When a fault manifests itself, the MCP (Maintenance Planning and Control) team schedules as soon as possible the corrective maintenance for repair or replacement, which results in a low number of acquired signals of fault conditions. In most conventional AI models, unbalanced datasets are associated with many issues, like for example low capability of prediction the least

---

represented class, as the model tends to learn more about the condition that is most presented to it;

(ii) *High cost to obtain and label data*: DL models are nowadays the standard given their performance and flexibility. However, for successful training, such models require great amount of data, which is time consuming to obtain or not always feasible. Data acquisition (Jan et al., 2023) is a challenging task and experts are typically required for data labeling: although the vast majority of faults have their characteristic behaviors known, the analysis requires specific knowledge that is not always available within the team. While *transfer learning* (Gentner et al., 2021; Li, Zhang, Qin, & Estupinan, 2020) approaches can overcome such weaknesses, the vast majority of approaches are based on obtaining at least the Source Domain with real and labeled signals, which does not fully solve the aforementioned problems.

In order to reduce the need for labeled real data, a few recent studies propose the use of synthetically generated data from numerical and physical models (Gecgel et al., 2021; Handikherkar & Phalle, 2021; Khan et al., 2021; Sobie et al., 2018). Despite the satisfactory results obtained, the approaches are limited in terms of adjustment of boundary conditions, element properties, definition of solvers, model complexity, and application to only one type of failure (in this case, bearing fault), which can limit large-scale industrial applications.

Similarly to the problem of obtaining a complete and labeled training set, the adoption of AI models in the industry comes up against its interpretability (Li et al., 2019). Called Explainable Artificial Intelligence (XAI), the area has been showing great interest by researchers, and is identified as one of the solutions to bridge the gap between AI researches and engineering applications (Lei et al., 2020; Zhao et al., 2022). Many models are considered as black boxes, that is, the user responsible for receiving the diagnosis does not know how the model reached the final conclusion. By not knowing exactly how or which features of the signal the model was based on for decision making, the reliability of the diagnosis is compromised, implying even the non-use of the model. To solve this problem, studies have been developed to explain why the model obtained the final classification. Among the techniques that provide visual explanation, the state-of-the-art technique in post-hoc methods is Gradient-weighted Class Activation Mapping (Grad-CAM). As faults in IFD using vibration are generally identified through a visual analysis of the signal in the frequency domain, it is interesting to provide a heatmap overlaid on the input signal, identifying the most relevant frequencies for the classification. Despite the wide application in other areas, at present, Grad-CAM algorithm is seldom used in AI models for fault diagnosis (Feng et al., 2020; Lin & Jhang, 2021; Yu et al., 2022).

To overcome the problems presented, a new interpretable approach for classifying faults in rotating machinery based on transfer learning from augmented synthetic data to real rotating machinery is proposed, namely FaultD-XAI (Fault Diagnosis using eXplainable AI). The approach starts from the concept where, transfer learning models focus on storing knowledge gained while solving one problem and applying it to a different but related problem. However, instead of working on fine-tuning, or retraining the model, the approach focuses on synthetic creation of the source domain: in this way the trained model learns features, which allow the transfer knowledge through the shared features of the source and target domain.

In FaultD-XAI, we propose to train the AI model with synthetic signals. The signals are generated from the knowledge about different faults in rotating machinery combined with the original signal of the machine under analysis, avoiding the use of complex mechanical models. The original signal is used only as a reference to adjust the synthetic signals, not requiring samples of all possible faults, which solves the problem of lack of labeled data in training set and enables engineering applications. To increase the variability of the training dataset, and consequently the robustness of the AI model, signals are generated using data augmentation techniques. To avoid the need to perform featuring engineering, reduce the computational cost, and

enable the model's interpretability through the use of Grad-CAM, 1D Convolutional Neural Network (1D CNN) is used. Vibration signals in frequency domain are used, bringing the analysis as close as possible to that performed by experts. Due to the importance of rotors, bearings and gears for rotating machinery and especially their respective faults, three datasets are used for validation of the methodology, each one being related to a respective component and its possible faults.

The use of synthetic data makes it possible to apply the AI model in different rotating machinery even without having real signals in all operating conditions. Such ability can be seen as a transfer learning approach, since the proposal can be used in any type of machine, just by modifying the reference signal. In addition, the model is trained with a dataset belonging to the source domain, which, in turn, presents a difference from the signals present in the target domain. The data augmentation of the data allows the model to be implemented quickly, since few collections are enough to generate the training data. Finally, the interpretability of the FaultD-XAI allows the user to have reliability and confirm the diagnosis, enabling its implementation, and even, for the person responsible for the development of the model, to verify the coherence of the learning during the training phase, allowing adjustments if necessary.

Although some of the processes used here (such as signal generation, augmentation and signal processing) can be considered mature, to the best of our knowledge, they have never been modified and integrated in such a way to overcome existing problems in the industry, such as the impossibility of obtaining real data of all kinds of faults, lack of transfer learning and interpretability and the need to always have a specialist to carry out the analyzes (implying high labor costs and difficulty in having this type of professional). Therefore, to the best of our knowledge, no studies have been published presenting an approach that overcomes these major problems in the industry.

In summary, three major problems hinder the application of AI models in industrial environments, being the motivation for the research: (i) impossibility or long term to obtain a sample of all operational conditions (since failures rarely happen), (ii) high cost of experts to label all acquired data (in a supervised scenario) and (iii) lack of interpretability of models (black boxes). And to overcome the principal problems that motivate the work, the main contributions of the proposed approach are: (i) a new transfer learning classification approach based on a synthetic dataset, without the need to have signals of real fault conditions; (ii) a post-explain approach to the fault diagnosis process, which allows interpreting the way in which the final result was obtained by the model, supporting decision-making which, in turn, is also a new contribution to the recent area of study: XAI in fault diagnosis; (iii) a generic and simple way to generate synthetic fault data for training, based on the knowledge available in vibration analysis, without the need for complex models; (iv) possibility of generating varied training datasets of different sizes (data augmentation-targeting deep learning applications); (v) new dataset, publicly available at the link: https://data.mendeley.com/datasets/zx8pfhdtnb/1, to study failures such as: unbalance, misalignment and looseness; (vi) possibility of application in different types of faults; (vii) faster deployment of the model in production; (viii) industrial application on real world datasets.

The remainder of this paper starts with a brief explanation about the 1D CNN and Grad-CAM. The proposed approach is presented in Section 3. Experimental procedure is shown in Section 4. Results and discussion are given in Section 5. Finally, Section 6 concludes this paper by drawing conclusions and discussing potential future works.

2. Related works and background

Kumar and Hati (2020) presented a review of the main Machine Learning (ML) and Deep Learning (DL) techniques applied in the monitoring of induction motors, aiming to detect faults such as: broken bars, bearings failures, stator failures and eccentricity. Liu et al. (2018) also present a review on artificial intelligence for fault detection in

rotating machinery, in which more than 100 cited references refer mostly to fault classification (fault diagnosis). More recently, (Lei et al., 2020) presented a comprehensive review with more than 400 citations, focused on Artificial Intelligence (AI) applications for fault diagnosis, providing a historical overview, in addition to current developments and future prospects. Further details on the state-of-the-art in Artificial Intelligence (AI) for fault diagnosis in rotating machinery can be found in Duan et al. (2018), Hoang and Kang (2019), Khan and Yairi (2018), Saufi et al. (2019), Zhao et al. (2019).

Lei et al. (2020), Zhuang et al. (2021) present some approaches developed with a focus on transfer learning, showing that transfer-learning theories can alleviated the lacking labeled samples and enhancing the applications of Intelligent Fault Diagnosis (IFD) in industry, in such works several approaches are considered, namely: feature-based approaches (Yang et al., 2019), generative adversarial network (GAN) based approaches (Guo et al., 2019), instance-based approaches (Shen et al., 2015), zero-shot learning (Xu et al., 2022) and domain adaptation (An et al., 2023).

Data augmentation is a widely used practice in ML when it is desired to increase the relevance of the dataset under study, mainly for image classification, natural language understanding, semantic segmentation and also in fault diagnosis (Alves et al., 2020; Khan et al., 2021; Li, Zhang, Ding, & Sun, 2020). In addition to increasing the amount of data for training, it is possible to insert small variations that, do not change the general context of the sample (although they exist). Some popular techniques are: flip, rotation, scale, crop, translation, gaussian noise. This makes the model able to correctly classify the sample, even if it presents noise, translations, different size, lighting etc. Performing augmentation can prevent the model from learning irrelevant patterns and consequently improve overall performance.

Just as the image of an object can be obtained from different perspectives by varying the context, but not the object under analysis, the same can be observed in the vibration signal of a rotating machinery. In an industrial plant, machines are subject to process variations, human interference, disturbances from nearby machines that can introduce extra information into the signal. Therefore, a robust classification model should be irrelevant to such variations, focusing only on identifying the operating condition of the asset under analysis.

In addition to improving data variability, data augmentation contributes to increasing data in the training set, enabling the implementation of more complex and deep learning models.

Related to synthetic data generation, stands out the methodology presented in Sobie et al. (2018); in the study, a simulation-driven machine learning for bearing classification is proposed where training signals were generated from high resolution simulations in Siemens LMS Imagine. Lab Amesim simulation software using the one-dimensional 3-DOF model.

2.1. 1D convolutional neural network (1D CNN)

Convolutional Neural Networks (CNNs) have become the de facto standard for Computer Vision and Machine Learning. CNNs are able to extract features automatically from the 2D data, avoids time-consuming analysis and signal processing that need to be done when working with other techniques, such as those shown in Brito et al. (2021).

In the context of vibration analysis, 2D CNNs have been widely used (Appana et al., 2017; Ding & He, 2017; Lu et al., 2017), however a more natural approach to deal with time-series data, are 1D CNNs (Abdeljaber et al., 2017; Avci et al., 2018; Ince et al., 2016). In the area of fault detection in rotating machines, some studies involving 1D CNN are Chen et al. (2019), Chuya-Sumba et al. (2022), Wang et al. (2021).

Among the main advantages of using 1D CNNs, the following stand out: (i) significant computational cost reduction w.r.t. 2D CNNs; (ii) more compact networks (with 1–2 hidden CNN layers and configurations with networks having less than 10 K parameters (Kiranyaz et al., 2021)); (iii) possibility of training on any CPU; (iv) possibility of being used for real-time and low-cost applications due to its low computational cost.

For more details on 1D CNN we refer the interested readers to Kiranyaz et al. (2019).

2.2. Gradient-weighted class activation mapping (grad-CAM)

Grad-CAM (Selvaraju et al., 2017) is an improvement on traditional CAM (Zhou et al., 2016). Studies show the use of the method in the monitoring of rotating machinery, as in Lin and Jhang (2021) that used 1D CNN and 1D LeNet-5[1] (I1DLeNet) to prove the efficiency of the proposed method SBDS (smart bearing diagnosis system), which uses a gradient diagnosis – weighted class activation mapping (Grad-CAM) – based convolutional neuro-fuzzy network (GC-CNFN). Proving that the method is not only capable of correctly classifying bearing faults, but also helping the user to understand the result.

Kim and Kim (2020) used Grad-CAM and Acoustic Emission Signals to detect bearing faults, again providing the user interpretability regarding the frequencies used by the model to obtain the result. Feng et al. (2020) applied the methodology in the monitoring of bearings and validated it in a gear fault dataset, showing that it is possible to use the methodology to identify the most attentive part of the model in relation to each type of fault. Yu et al. (2022) used Grad-CAM and eigenvector-based class activation map (Eigen-CAM) to interpret the ResNet06 (a popular CNN architecture (He et al., 2016) in 4 databases, 3 bearing and 1 gearbox dataset. In order to interpret the effectiveness of the method, Grad-CAM is applied to localize the regions in the input that contribute the most to the network's prediction. The proposed method is validated by a motor bearing dataset and an industrial hydro turbine dataset (Wu et al., 2021).

Grad-CAM uses the gradients of any target concept flowing into the final convolutional layer to produce a coarse localization map highlighting the important regions in the image for predicting the concept (Selvaraju et al., 2017). The signal is forward propagated by the CNN part of the model and then processed to obtain the raw score for the category. Gradients are reset for all classes except the class under analysis, set to 1. The signal is then backpropagated to the rectified convolutional feature maps of interest, which is combined to generate the heatmap. The highlighted points (i.e., the highest gradient value) are the most relevant regions that the model uses to make the decision. For more details about Grad-CAM please refer to Selvaraju et al. (2017).

3. Proposed approach: FaultD-XAI

The proposed approach is presented in Fig. 1, and consists of six parts: (i) Data Acquisition; (ii) Signal Generation; (iii) Data Augmentation; (iv) Signal Processing; (v) Fault Diagnosis; vi) Explainable Artificial Intelligence (XAI). Initially the real signals (x_r) are collected from the machine. To create different operating and fault conditions, synthetic signals (x_s) imitating fault characteristics are generated and then combined with real signals (x_r). To reduce the amount of original data needed for training the model, the signals are augmented ($x_r x_s + x_a$). The fourth step consists of the last phase of data preparation, where the signals are converted from the time domain to frequency domain. After pre-processing the data, the model is created and trained, enabling the diagnosis of the machine's condition (fault diagnosis). Finally, the model is explained, identifying the most relevant frequencies used for classification, supporting the user's decision making. A more detailed description of the methodology is presented in Algorithm 1.

To the best of our knowledge, FaultD-XAI is the first approach that combined synthetic data with data augmentation to avoid the need of real fault data by exploiting XAI and transfer learning for fault diagnosis in rotating machinery.

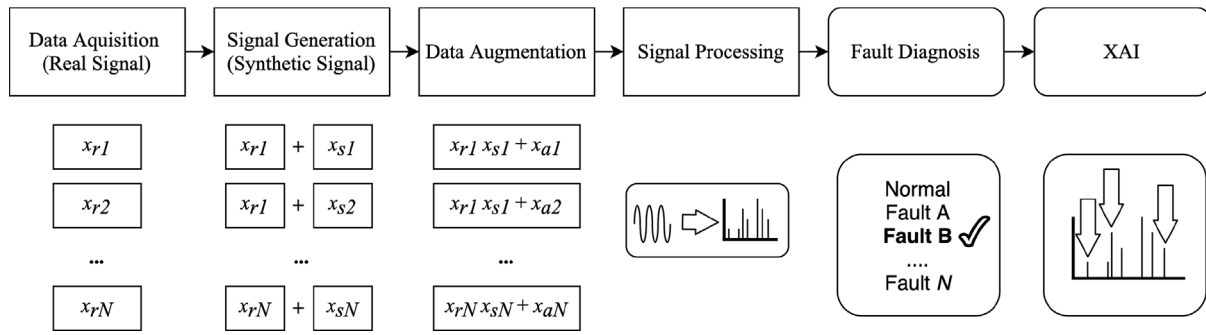---

[1] LeNet is classic CNN architecture (Lecun et al., 1998).

Fig. 1. FaultD-XAI: General framework of the proposed methodology.

Algorithm 1 Pseudo-Code

```
 1: procedure FaultD-XAI
 2:     Type: fault classification / XAI analysis
 3:     Input: new sample
 4:     Output: fault classification and interpretability
 5:
 6:     initialization (input data) ← def(BPFO, BPFI, BSF, FTF, Rotation Frequency)
 7:     load real signal ← load(real signals for training set (x_real))
 8:     generate synthetic fault signals ← def(fault characteristics (f_fault): 1×fr, 2×fr, BPFO, GMF ...)
 9:     generate synthetic fault signals ← def(set amplitude (A) based on real signal (aprox. 3dB))
10:     generate synthetic fault signals ← def(generate different signals for each fault added)
11:     for i in range(f_fault) do
12:         e.g. : x_syn = A × sin(2π × f_fault × t)
13:     data augmentation (add variation) ← def(gaussian noise, signal translation ... (z_noise))
14:     for i in range(x_syn) do
15:         x_aug = x_syn + z_noise
16:         repeat for each type of variation, n times defined
17:     signal processing ← pxx = pwelch(x_aug)
18:     signal processing ← normZ(pxx)
19:     fault diagnosis ← training(training set: pxx_training)
20:     fault diagnosis ← test(test sample)
21:     print('The fault is located in: ', predicted value)
22:     XAI analysis ← load(test sample)
23:     plot(heatmap: grad cam predicted values)
```

3.1. Data acquisition

Among the sensors used for monitoring rotating machinery, the vibration-based diagnostic method is the most popular and researched. The interest is justified by the fact that the vibration signals directly represent the dynamic behavior of the equipment (Ciabattoni et al., 2018; Dalvand et al., 2017; Samuel & Pines, 2005) and are a non-invasive technique (Brito et al., 2022). Therefore, vibration signals were used.

The real signal is fundamental to generate the synthetic dataset, enabling transfer learning between the source domain (synthetic signals) and target domain (real fault signals). Through the real signal of the machine under analysis, it is possible to generate a dataset that presents similar characteristics, favoring the learning of the model. It is worth mentioning that the real signal used is the current operating condition of the machine, not requiring a sample of each condition.

It is noteworthy that, although various types of fault signals can be obtained using synthetic signals, it is necessary for engineers to collect a few real machine vibration signals, so that factors such as amplitude, background noise, influence of machine vibrations nearby, are adjusted to the real signal in the data augmentation process.

3.2. Signal generation

Engineering applications involve a major limitation regarding the data available for training the AI model. In the vast majority of cases, we do not have real and labeled samples of all possible operating conditions, since faults seldom happen. In applications where data is available, it takes a lot of time and cost with experts to label and analyze all signals. A strategy to overcome this problem is to create the training dataset synthetically, since the vast majority of faults in rotating machines have known characteristic patterns.

Different approaches can be used to model faults in rotating machinery, such as: Digital Twin, where through equations of motion, numerical models, optimization and support software, the machine is created virtually, and through modifications in its conditions it is possible to verify the signal response, identifying a fault or a new operational condition.

Despite the available approaches, modeling a rotating machinery is not always an easy task or a low computational cost, which can make some industrial applications unfeasible. It is known that many parameters are highly related to the machinery and cannot be randomly assigned, for example, the amplitudes of a signal, damping factors of a bearing repetitive transients etc. Therefore, we chose to model the faults, in a simple way, through the equation of a waveform, with a sine function, considering that the faults are mostly related to deterministic frequencies or characteristic behaviors in the signal. And in this way, model the fault according to its behavior analyzed by the vibration specialist in the frequency spectrum of the signal. The main goal is not to use complex dynamic models that take into account all the dynamic interactions of the equipment, which, despite being assertive, are equally complex and need to be developed machine by machine. On the other hand, using human knowledge about how the faults behave in the signal, and creating synthetic signals that replicate their behavior,

enabling the implementation of the system in an industrial environment quickly and efficiently.

In general, the most common faults found in the industry are related to bearing defects, classic problems such as unbalance, misalignment, looseness, as well as problems in gears, and for this reason they were chosen for analysis of the proposed approach. It is worth mentioning that the way the FaultD-XAI is proposed, other faults can be modeled and added, without changing its functioning (e.g., cavitation, electrical defects, oil whirl and oil whip etc.) It is up to the specialist to determine the synthetic faults that will be created referring to the machine under analysis. It is also important to highlight that the methodology applies to stationary operating conditions, considering that the faults are created as a function of certain frequencies, which are mostly associated with the rotation frequency of the machine.

The representation of the oscillatory signal *x(t)* used to generate the faults (unbalance, misalignment, looseness and gear fault) is defined as:

$$x(t) = A \times sin(2\pi \times f \times t + \theta) \tag{1}$$

Where A indicates the peak amplitude of the signal, f frequency [Hz], t time vector [s] and θ phase [rad]. As accelerometers and no tachometer are used, the phase reference was not applied, although it can help in the identification of some faults.

The amplitudes were randomly generated so that the synthetic fault signal presented an approximate variation of at least 3 dB in the characteristic frequencies of the fault in relation to the normal signal. Since, according to ISO (2009) and the experience of experts in vibration analysis, a variation of less than 3 dB does not represent a significant change in the behavior of the machine in terms of predictive maintenance or failure. The amplitudes are randomly generated to represent different stages of fault, and improve the robustness of the AI model (greater variability in the training group). It is worth mentioning that, depending on the fault, minor variations can be significant, and it is up to the specialist to determine when assembling the model.

The unbalance occurs predominantly at 1 x fr (Rotation Frequency), so a signal was generated with a frequency equal to the rotation frequency. It is worth mentioning again that the phase is not being taken into account to differentiate the faults, only the characteristic frequencies of the signal.

Misalignment presents high vibration values at 1x, 2x and 3x rpm, and can be angular, parallel or mixed (when the two are combined). In that case, Eq. (1) was used three times, being the final signal the sum of each equation with a respective frequency (1x, 2x and 3x rpm).

The looseness caused by loose pillow block bolts, cracks in frame structure or in bearing pedestal, or by improper fit between component parts, causes harmonics and sub-harmonics of the rotation frequency. The signal will present multiples of the rotation frequency as: 1x, 2x, 3x, ...Nx and sub-harmonics: 0.5x, 1.5x, 2.5x, ...$N/2x$, where N is the multiple of the rotation. As well as the misalignment, the final signal was composed of the sum of harmonics and sub-harmonics.

Gear faults are usually related to Gear Mesh Frequency (GMF) and its harmonics. GMF is the product of the number of teeth on the gear (n_{teeth}) multiplied by the running speed of the gear (g_{fr}), described as:

$$GMF = n_{teeth} \times g_{fr} \tag{2}$$

Defects such as wear, misalignment or eccentricity and backlash in gears may not only cause a significant increase in 1x GMF. In many cases, there is an increase in the amplitude of its harmonics as 2x GMF and 3x GMF. Such defects not only cause variation in the GMF and its harmonics, but in general, also present sidebands of the rotating speed of the defective gear. Therefore, the faults were simulated using Eq. (2) where, the frequency for each signal is given by GMF and its harmonics (1x, 2x, 3x GMF) and by the sidebands (GMF +/- rotating speed of the defective gear). The final signal was composed by the sum of each signal generated.

The faults in rolling element bearings are related to their components: inner race, balls or rollers, cage and outer race. Deterioration will cause characteristic frequencies in the signal allowing fault identification. The four frequencies are: (i) BPFO (Ball Pass Frequency Outer); (ii) BPFI (Ball Pass Frequency Inner); BSF (Ball Spin Frequency); (iv) FTF (Fundamental Train Frequency). The calculation basically takes into account: the number of elements, internal and external diameter, contact angle, and its equation and value can be easily found in vibration analysis software, supplier catalogs and existing literature.

High frequency resonances between the bearing and the response transducer are excited when the rolling elements strike a local fault on the outer or inner race, or a fault on a rolling element strikes the outer or inner race (Randall & Antoni, 2011). The impact frequency depends directly on the machine components and their natural frequencies, however, as the interest is to evaluate the characteristic fault frequencies, this value can be arbitrarily defined in the signal modeling to obtain the value of the characteristic frequencies of bearing fault: BPFO, BPFI, FTF and BSF. The impact signal can be modeled using Eq. (1) where A indicates the peak amplitude of the signal, f impact frequency [Hz], t time vector of impact [s]. To make the impact periodic, it is convoluted with a Comb function, where the non-null values (periodic peaks) correspond to the fault frequencies of the elements. The Comb is a sum of time shifted Dirac Delta, that is, it is defined to be zero at alternate time points.

After creating the synthetic faults, the original signal is added to represent the dynamic behavior of the machine, allowing the transfer knowledge. Due to the types of faults under study, at the end of the process, each original signal will result in 7 signals, being: normal condition, outer race fault (BPFO), inner race fault (BPFI), unbalanced, misaligned, mechanical looseness and gear fault (due to their common usage, the terms BPFO and BPFI will also be used to identify the type of fault under analysis, that is, fault in the outer and inner races, respectively.) Although the inner race fault (BPFI) was not present in any of the cases studied, it was introduced to increase the complexity of the training and also validating the FaultD-XAI.

3.3. Data augmentation

For the proposed approach, each synthetic signal is augmented, randomly varying the method variables. The methods presented in Li, Zhang, Ding, and Sun (2020) were used. It is worth mentioning that other methodologies can also be used. The proposed methods are:

(i) Gaussian Noise: apply random noise to the raw signal, as follows:

$$\bar{x} = x + \alpha_{gauss} \times G, \ G \sim N(0,1) \tag{3}$$

where $\alpha_{gauss} > 0$ is the Gaussian noise coefficient, $\bar{x}$ is the augmented data, x is the raw signal and G the gaussian noise.

(ii) Masking Noise: an alternative way to apply random noise, where a fraction α_{mask} of the elements in x are randomly selected and set to 0.

(iii) Signal Translation: the signal is randomly shifted forward and backward, with the data gap resulting from the translation complete with zeros, maintaining the original size of the signal.

(iv) Amplitude Shifting: Moderate variations in signal amplitude are common, and do not significantly alter the machine's operating state. Therefore, performing modifications of this nature tends to enrich the dataset. Amplitude shifting is implemented through a scaling factor (α_{scal}):

$$\bar{x} = \alpha_{scal} \times x \tag{4}$$

(v) Time Stretching: The signal is stretched along the time axis by a stretching factor ($\alpha_{stre} > 0$). A sample with N_{aug} values in the range of [1, α_{stre}] if $\alpha_{stre} \geq 1$ or [α_{stre},1] is obtained. The sample is centered and stretching is applied. Points outside the dimensions are clipped out and gaps filled with zeros.

At the end of the data augmentation process, each signal will generate five new signals, and the process can be repeated n times. In

summary, the entire process of creating synthetic signals and augmentation from one original signal will result in 35 new signals (7 synthetic signal * 5 augmented signal). The number of times (q_{aug}) the procedure needs to be executed to obtain a total of n signals (n_{total}) in the training group can be calculated as follows:

$$q_{aug} = \frac{n_{total}}{7 \times 5 \times n_r} \tag{5}$$

where, n_{total} is the desired amount of samples in the training set, 7 is the amount of synthetic signals, 5 is the amount of augmented signals and n_r the amount of real signals used. When the relationship does not result in an integer, the value is rounded up to the nearest integer, and the excess is randomly selected and discarded.

3.4. Signal processing

DL models are able to extract features from raw signals and learn characteristic behaviors without the need for feature engineering as in classic ML models.

Although the raw signal (time domain) presents relevant information, due to the amount of frequencies present, the analysis often becomes complex. Therefore, it is common to perform analysis in the frequency domain. As one of the goals of the work is to provide interpretability of the result, the input signal was used in the frequency domain instead of the time domain.

The signal was normalized using Z-score normalization, and finally, cut to reduce the number of points, and consequently the computational cost of the model. For the cut, the maximum frequencies at which the fault could manifest and their harmonics were analyzed, so according to the monitored asset, the size of the signal used may vary. As the dataset under analysis present constant speed, it is not necessary to perform order tracking or other method as a pre-processing step to remove the effect of shaft speed variations.

3.5. Fault diagnosis

CNN are present in several applications due to their main characteristics: (i) possibility to learn features directly from raw signal in training; (ii) are immune to small transformations/disturbances in the input data; (iii) can adapt to different sizes (the model can be trained to receive different sizes of input data, but the architecture of the model is pre-defined during training); (iv) excellent computational cost compared to conventional fully-connected Multi-Layer Perceptrons (MLP) networks (CNN neurons are sparsely-connected with tied weights Kiranyaz et al., 2021).

Since AlexNet's proposal (Krizhevsky et al., 2012), when it achieved 16.4% error rate in the ImageNet benchmark dataset (this was about 10% lower than the second top method), several deep 2D CNNs have been used. However, there are certain drawbacks and limitations of using such deep CNNs: (i) high computational cost and special hardware for training; (ii) requires massive size dataset for training to achieve reasonable generalization capability (Kiranyaz et al., 2021). Factors that are not always available when it comes to industrial applications involving vibration signals. In addition, vibration signals are in 1D, which would make it necessary to modify their dimension, further increasing the computational cost and need for pre-processing. On the other hand, 1D CNN makes it possible to use the vibration signal in its original dimension. It has low computational cost compared to 2D CNN and is well-suited for real-time and low-cost applications especially on mobile or hand-held devices (Kiranyaz et al., 2021). For these reasons, it was used in the study.

In this part, the synthetically generated and augmented signals are used to train the model (1D CNN) and adjust the hyperparameters. It is worth mentioning that no real fault signals are used in the training. The test group consists of other real machine signals under different operating conditions. After training, the model is ready to perform the classification.

3.6. Explainable artificial intelligence (XAI)

Explaining the results obtained by the AI model is essential to ensure its use in the industry, since it is associated with increased trust. Vibration signals are preferably analyzed in the frequency domain in order to facilitate the correlation between the highlighted frequency/behavior and the fault pattern. In this way, applying a method that evaluates the points of relevance in the signal used for the classification, allows the end user to understand and validate the result obtained. For this purpose, the Gradient-weighted Class Activation Mapping (Grad-CAM) approach was used.

After classifying the sample by CNN 1D, the method is applied, returning a heatmap with the most relevant frequencies for the model. For time-series data, Grad-CAM computes the most important time steps for the classification decision of the network, calculated by propagating the gradients in the convolutional layers. The places where this gradient is large are exactly the places where the final score depends most on the data. In the case of the vibration signal in the frequency domain, the deterministic frequencies used for fault identification correspond. The heatmap obtained helps the specialist in decision making, allowing the validation of the classification determined by the model. When it comes to the development of the model, it allows analyzing the main reasons (main frequencies) that are leading to the error or success, enabling the adjustment in the development phase.

4. Experimental procedure

4.1. Data description

Three datasets were used to address different faults found in rotating machinery, two of which were publicly available (Cao et al., 2018; Qiu et al., 2006) and one developed by the author for the study. As shown in Lei (2017) even though the rotating machinery is diversified, some common essential rotating parts are: rotors, rolling element bearings, and gears (Brito et al., 2022). Therefore, the datasets were chosen and developed to address the main faults present in rotating machinery components, namely: defects in bearing and gearbox, misalignment, unbalance and mechanical looseness. The use of different datasets aims to validate the proposed methodology in different scenarios. As they are publicly available datasets and already explored in Brito et al. (2022), Case 1: Bearing Dataset and Case 2: Gearbox Dataset will be briefly discussed. For more details, please refer to Brito et al. (2022), Cao et al. (2018), Qiu et al. (2006). All datasets are vibration signals collected using accelerometers.

4.1.1. Case 1: Bearing dataset

The dataset (Qiu et al., 2006) is composed of remaining useful life (RUL) bearing tests, each file consisting of 20,480 points with the sampling rate set at 20 kHz. The rotation speed was kept constant at 2,000 rpm by an AC motor coupled to the shaft via rub belts. A radial load of 6,000 lb was applied to the shaft and bearing by a spring mechanism. Rexnord ZA-2115 double row bearings were installed on the shaft. As this is a run-to-failure test, no labels are available (the only information provided is the type of fault present at the end of each test). Therefore, the data were analyzed and labeled manually. For the study the bearing 01 of test 02 was used. The test presents 984 observations, with the first 531 observations labeled as normal and the last 453 as fault in the outer race (BPFO).

4.1.2. Case 2: Gearbox dataset

To evaluate faults in gears, the dataset (Cao et al., 2018) was used. A 32-tooth pinion and an 80-tooth gear were installed on the first stage input shaft. The second stage consists of a 48-tooth pinion and 64-tooth gear. The data were recorded using an accelerometer through a dSPACE DS1006 system, with sampling frequency of 20 kHz. Nine different gear conditions were introduced to the pinion on the input shaft, including

Fig. 2. Bench test.

healthy condition, missing tooth, root crack, spalling, and chipping tip with five different levels of severity. For each gear condition, 104 observations were collected resulting in a total of 936 observations with sampling frequency of 20 kHz. For the test, the samples were considered as: 104 normal and 832 gear fault.

4.1.3. Case 3: Mechanical faults dataset

The last dataset was developed by the authors, aiming to address classic faults in rotating machines such as: unbalance, misalignment and mechanical looseness, in addition to the normal operating condition.

The faults were introduced on a test bench, Fig. 2, being: motor, frequency inverter, bearing house, two bearings, two pulleys, belt and rotor (disc).

- Motor: Three-phase induction motor, B56 B4, Manufacturer Eberle, nominal speed 1650 rpm, power 0.09 kW, voltage 220 V, current 0.70 A, bearings 6200 ZZ.

- Frequency Inverter: Vector Inverter, model CFW300, manufacturer WEG.

- Driven Pulley (Disc Side): model A80, external diameter 80 mm, internal diameter 53 mm.

- Motor Pulley (Motor Side): model A60, external diameter 60 mm, internal diameter 37 mm.

- Belt: model V - A23, top width 12.7 mm, outer perimeter 630 mm, inner perimeter 580 mm.

- Rotor (Disc): diameter 149 mm, 36 holes, thickness 6.5 mm.

- Bearings 1205, bearing house SN505, bushing H305.

For the acquisition of the signals were used: 04 accelerometers PCB 352C33 (2 in each bearing), mounted in the vertical and horizontal positions (y- and x-axes), 01 Power Supply PCB 482A20, 01 acquisition board Hi-Speed USB Carrier, NI USB-9162. A Python language program was developed to collect the data. All accelerometer sensitivities were adjusted according to the calibration chart. The rotation was kept constant with a value measured on the shaft of approximately 1238 rpm.

20 tests were performed, 5 for each condition (normal, unbalance, misalignment and looseness). Each test consists of 4 sets of 420 signals collected continuously, each file consisting of 25,000 points with the sampling rate set at 25 kHz (420 signals per accelerometer), resulting in the end of all tests, in a total of 8400 signals per accelerometer.

The sequence of tests was randomly defined. Before starting a test, the bench was dismantled and returned to normal operating condition,

to later introduce the fault. The experimental procedure allows variations to occur, making the tests closer to industrial reality. All data used in this study will be made available at the link: https://data.mendeley.com/datasets/zx8pfhdtnb/1.

To increase the randomness of the tests in the fault classification (AI), and to validate the robustness of the model, only 3 tests per condition were used in the analyses, resulting in 12 tests in total (5040 signals). Since the fault classification was performed 10 times, at each new run of the model, 3 tests were randomly selected for each class of fault.

As it is a small bench test, the measurement points are close, and in order to reduce the computational cost, only the signals from the horizontal position of the accelerometer present in the bearing (near the pulley) were used in the analyses.

4.2. Hyperparameter tuning and evaluation metrics

The hyperparameters were adjusted based on the training set to obtain the best performance and stability in the model. The signals used to perform the grid search are different from those used for model training, in order to avoid problems such as overfitting. The cross-validation method was used, where the signals were randomly chosen between the training and validation group, being 90% for training and 10% for validation. It is worth mentioning that the training group is composed only of synthetic signals, and therefore does not contain any real fault signals. The test set is composed of real signals (not used in the training group).

The model was composed of a 1D CNN layer, followed by a dropout layer for regularization, then a pooling layer. The objective of using only 1 layer is to reduce the complexity and computational cost of the model. As CNN has high learning capacity, the dropout layer was used to slow down the learning process and hopefully avoid overfitting. Pooling learning reduces the amount of learned features by consolidating them and keeping only the essential ones, which prevents the network from learning features uncorrelated with the condition.

The learned features are flattened into a single long vector and passed to the fully connected layer before the output layer used to make a prediction. The fully connected layer provides a buffer between learned features and the output with the intention of interpreting learned features before making a prediction.

For the model, 32 parallel feature maps and kernel size of 5 were used. ReLu activation function and Adam optimizer were also used. To automate the training, the model was trained with a variable number of epochs, based on the Early Stopping technique, with a patience of 8 and a variation of 0.001 in relation to the validation accuracy. Batch size of 32 samples was used. Categorical cross entropy loss function and Softmax (activation function) were used as it is a multi-class classification problem.

The model was trained to classify 7 conditions, regardless of whether the case under analysis presents the fault or not, being: normal condition, BPFO, BPFI, unbalance, misalignment, looseness and gear fault. This simulates a condition where different faults may appear on rotating machinery, and the model will be prepared to identify them. In addition, it increases the complexity of the analyses.

Neural networks are stochastic, so at each new training a different model will be obtained, even if using the same dataset. It is worth mentioning that, in addition to the stochasticity of the networks, there is also the randomness in the generation and augmentation of the signals. To ensure that the model is stable and robust to variations in both network training and signal generation, the analyzes were repeated 10 times. At each analysis, new synthetic signals were randomly generated and the model trained. As in all dataset, the number of samples per condition were approximately equal, accuracy was used. For Case 2 where the test data are imbalanced, despite the use of accuracy, the confusion matrix is shown. To measure the variability after the 10 analyses, the standard deviation was calculated. The tests were performed using 2.2 GHz Intel Core i7 Dual-Core, 8 GB 1600 MHz DDR3, Intel HD Graphics 6000 1536 MB.
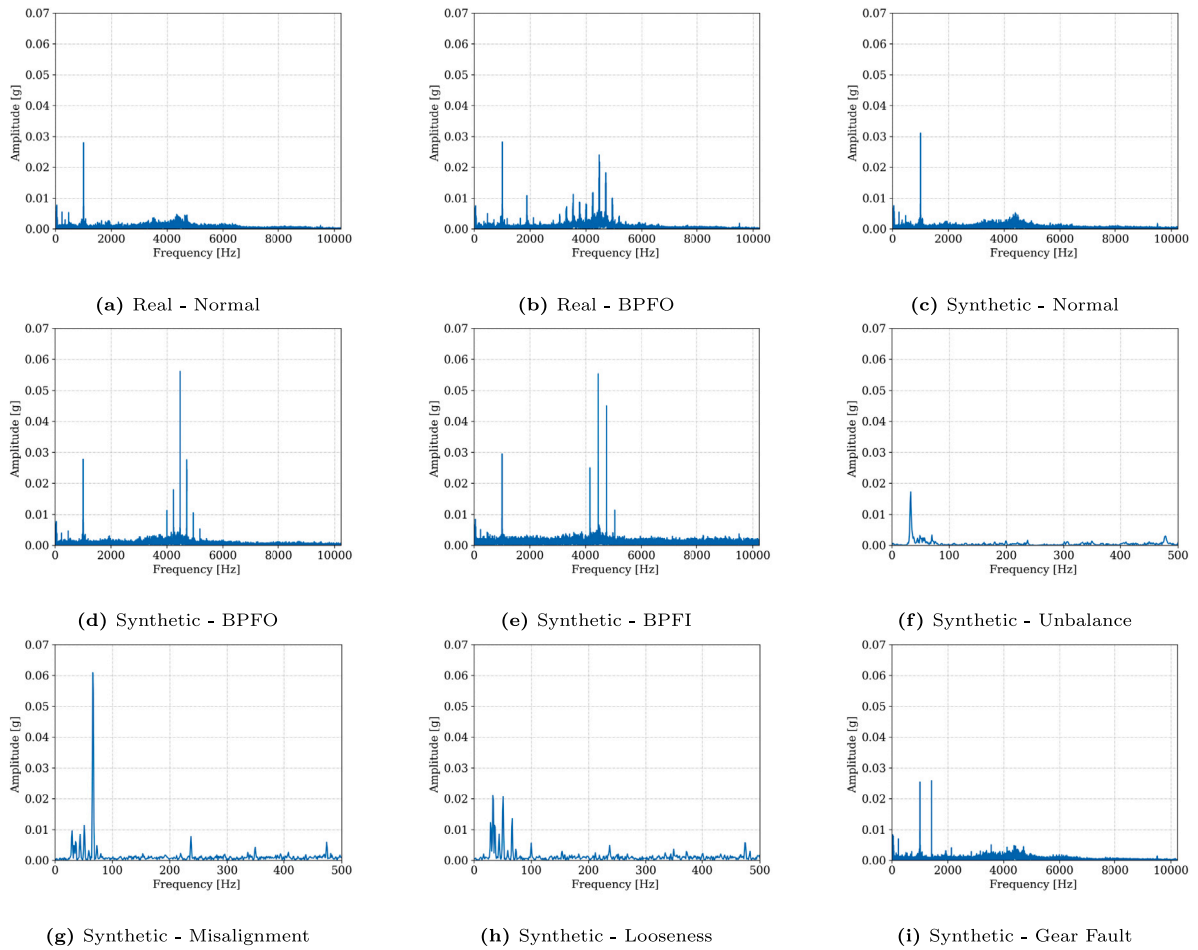
Fig. 3. Examples of vibration signals for Case 1, real and synthetic.

4.3. Analysis approaches

The success of an AI model is directly related to the quality and quantity of training data. On the other hand, a large amount of data implies acquisition costs, as well as a low quality implies a reduction in the model's accuracy. Being able to reduce the amount of real data needed, and generate data varied enough to make the model robust, is a solution to enable the implementation of AI in the industrial application.

Besides the main objective of verifying the possibility of using a model trained with synthetic and augmented data to evaluate real signals and their explainability, other analyzes were performed.

First, the augmented data are analyzed, verifying if the behavior of the synthetic signal in the frequency domain, matches the available knowledge about fault detection in rotating machinery using vibration analysis. Subsequently, in order to reduce the amount of real data needed to generate the synthetic data, the minimum viable amount of data was analyzed to obtain a stable and robust model. This approach can be seen as an analogy of the study area called Few-Shot Learning, where few samples are used to train the model. Another analysis performed is in relation to the total amount of real samples needed in the training group to ensure the stability and robustness of the model. In addition, the possibility of working with 100% synthetic samples was analyzed, that is, where synthetic data are generated without dependence on the real sample. For comparison between training forms, supervised training was performed, using only real signals, which were divided into 70% for training and 30% for testing. Finally, the data obtained through XAI are analyzed to verify if the features learned by the model are capable of providing relevant information to the expert for decision making.

5. Results and discussion

5.1. Data exploration

In this section the synthetic samples for Case 1–3 are analyzed, discussed and compared with the real samples. One sample per condition from the training database was randomly selected. Due to the randomness of FaultD-XAI, variations in signals may occur, being important to observe whether the deterministic characteristics of the conditions remain similar.

As the objective is only to visualize the signal characteristic, df = 1 hz was used, with the exception of Case 2, where df = 5.5 hz due to data limitation. For BPFO, BPFI and Gear Fault conditions, the full signal is presented to highlight the fault characteristic at high frequency. In Unbalance, Misalignment, Looseness, a zoom was performed in the range from 0 to 500 hz to facilitate the visualization of the characteristic frequencies. For the normal condition, the complete signal is displayed.

5.1.1. Case 1 - bearing fault

In Fig. 3 the original signals for the two conditions (Normal and BPFO) and the synthetic signals for the seven created conditions (Normal, BPFO, BPFI, Unbalance, Misalignment, Looseness and Gear Fault) from Case 1 are presented.

The real normal condition signal, Fig. 3(a) does not present any frequency related to faults, showing only the characteristic behavior of the machine. On the other hand, in Fig. 3(b) there is the appearance of frequencies in the region of 3000 to 6000 Hz, related to excitation of the natural frequencies, caused by the fault in the outer race.
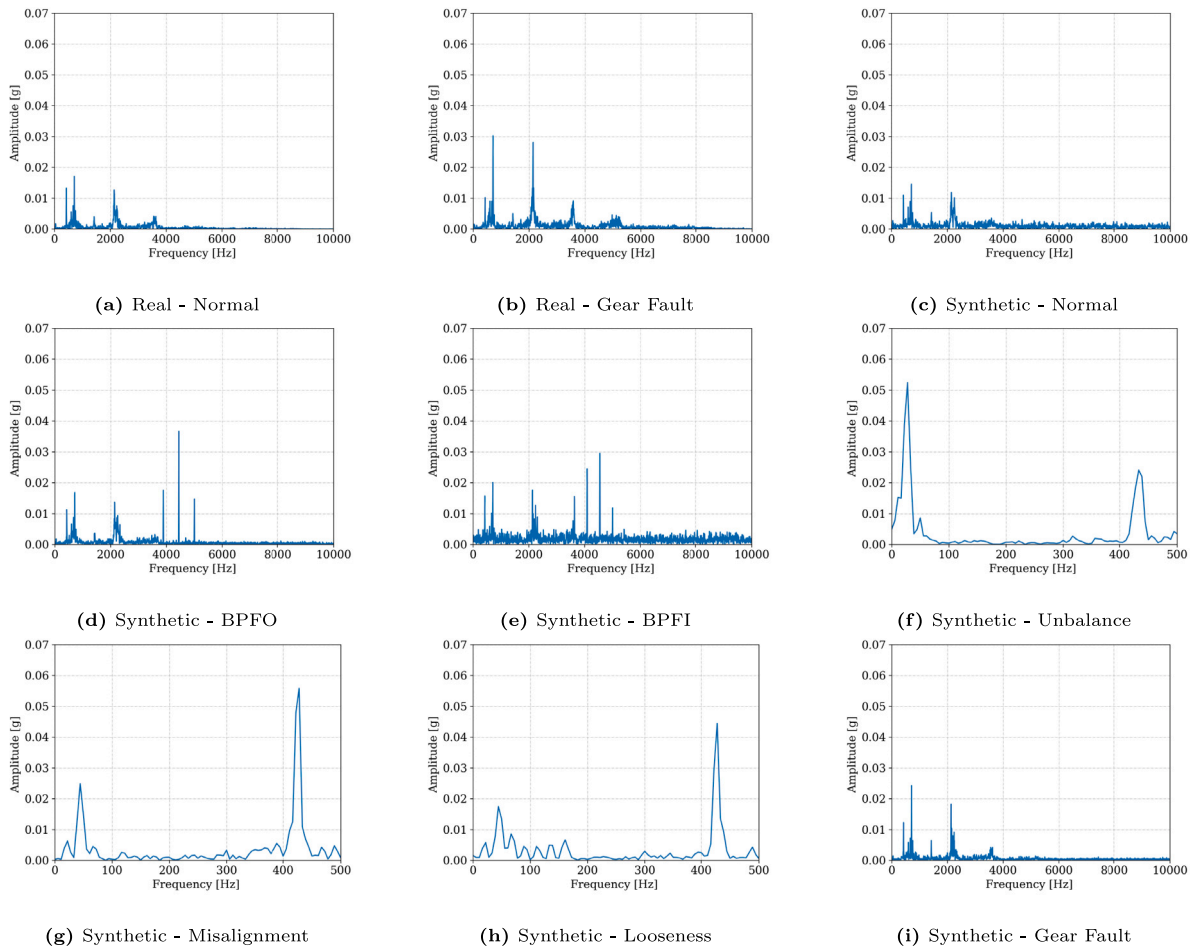
Fig. 4. Examples of vibration signals for Case 2, real and synthetic.

Analyzing the synthetic signals, it can be noted that for bearing faults, the impacts due to a small defect tend to excite the races natural frequencies (at high frequency) Fig. 3(d) e 3(e), being the fundamental frequency corresponding to the damaged element, BPFO = 236.4 hz and BPFI = 296.9 hz. For the unbalance, Fig. 3(f), there is a predominant increase in amplitude by 1 x fr (rotation frequency). In misalignment, there is an increase in 2 and 3 x fr, with emphasis on 2 x fr, Fig. 3(g). Finally, for the mechanical looseness, there is an increase in harmonics and the appearance of some sub-harmonics of fr, Fig. 3(h). Gear defect, Fig. 3(i) show an appearance of frequencies related to GMF, hypothetically placed at 711 hz and its harmonics, since this dataset does not have gears.

Comparing the synthetic and original (real) signals for the conditions present in the dataset (normal and BPFO), the similarity of the fundamental characteristics is noted, ensuring that the model learns the behavior of the signal in relation to each operating condition of the equipment.

5.1.2. Case 2 - gear fault

In Fig. 4 the original signals for the two conditions (Normal and Gear Fault) and the synthetic signals for the seven created conditions (Normal, BPFO, BPFI, Unbalance, Misalignment, Looseness and Gear Fault) from Case 2 are presented.

The real normal condition signal, Fig. 4(a) shows evidence of 1 and 2 x GMF, which is normal when monitoring gearboxes. On the other hand, the original faulty signal, Fig. 4(b) already presents, in addition to 1 and 2 x GMF, harmonics in 3 and 4 x GMF, and side bands, characterizing the appearance and progression of the fault.

Regarding the synthetic signals, the same analysis performed for Case 1 is valid for Fig. 4, with the exception of Fig. 4(i) which has GMF equal to 711 hz, being the real frequency of the gearbox. It can be noted that the behavior of the real and synthetic signals for each condition is similar, as proposed by the approach.

5.1.3. Case 3 - mechanical faults

In Fig. 5 the original signals for the four conditions (Normal, Unbalance, Misalignment and Looseness) and the synthetic signals for the seven created conditions (Normal, BPFO, BPFI, Unbalance, Misalignment, Looseness and Gear Fault) from Case 3 are presented.

Note, in Fig. 5, that even for the normal condition, there is evidence of highlighted frequencies up to 2000 hz, which may be related to lack of precision in the adjustment, assembly and lubrication of the bearings, improper adjustment (tensioning) and/or wear of the pulleys, noise generated by the frequency inverter etc. The condition was purposely chosen to simulate an industrial situation. It is known that, not all companies have enough specialized technicians to carry out an assembly and adjustment within the World Class Maintenance (WCM) standards. Also, there is a lack of specialized equipment to perform laser alignment, dynamic balancing etc. Moreover, due to the need for quick intervention to release the equipment for operation, maintenance cannot be performed with the desired precision. Because of these problems, the vibration signal can present a series of frequencies and/or noises, which make it difficult for the predictive maintenance specialist to analyze the main condition.

Even in a scenario where the signal is heavily polluted, technicians are usually able to perform an accurate diagnosis, based on field experience. Thus, it was decided to keep the bench in the presented
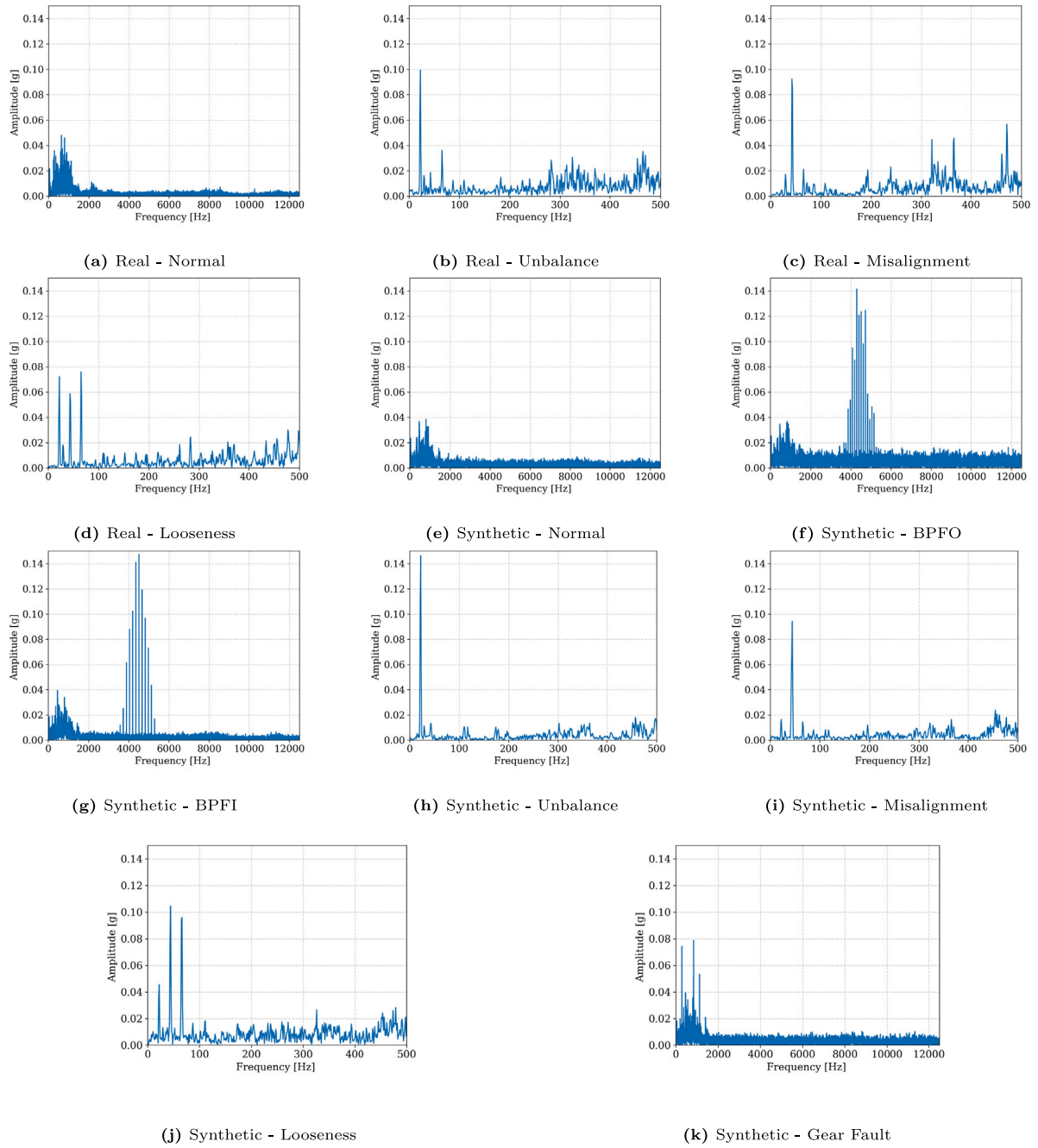
Fig. 5. Examples of vibration signals for Case 3, real and synthetic.

condition, to make the test closer to the industrial reality, instead of completely adjusting the bench to eliminate noise and other sources of excitation that are not related to the defects. Allowing, to assess whether the proposed methodology is capable of imitating human behavior, accurately identifying the operating conditions of the asset.

Again, analysis performed for Case 1 and 2 in relation to synthetic data is valid for Fig. 5, with the exception of Figs. 5(f) and 5(g) being the fundamental frequency corresponding to the damaged element, BPFO = 107.09 hz and BPFI = 155.7 hz.

As in the other cases, it can be noted that the described fault characteristics are evident in all the synthetically generated signals and in the original signals. This shows that the data augmentation process was able to increase the diversity of the dataset without losing its fundamental characteristics.

5.2. Number of total samples for training

To evaluate the influence of the total number of signals in the training group, 10 sets were created containing: 1050, 2100, 3150, 4200, 5250, 6300, 7350, 8400, 9450, 10500 signals.

As the objective was not to evaluate the efficiency of the model, but to compare the result obtained in relation to the amount of signals in the training group, 30 original signals were selected to perform the data augmentation step, and the results for accuracy and standard deviation are shown in Fig. 6.

Analyzing Fig. 6, it can be noted that there is no significant variation in relation to the amount of data used, with the exception of Case 3 in the set of 1050 signals. Such variation may have occurred due to the greater complexity of the signal present in the dataset, requiring
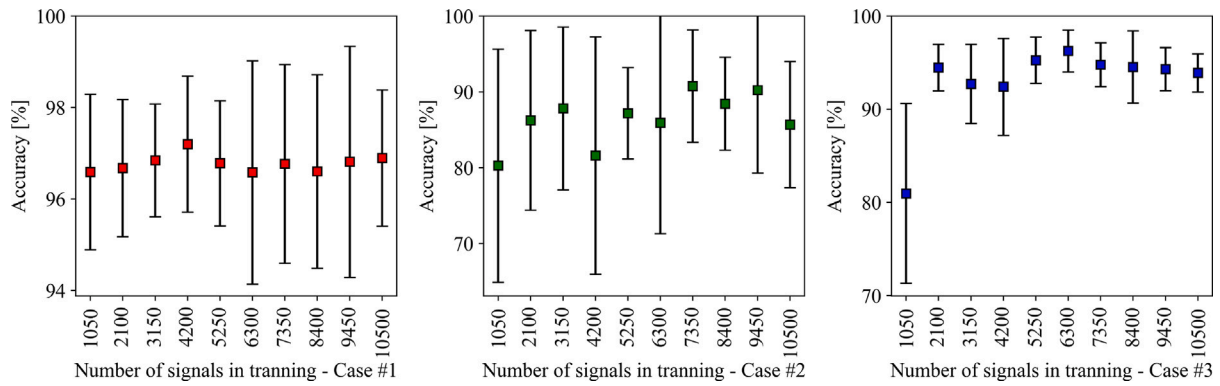
Fig. 6. Results for each case varying the total amount of signals in the training set.
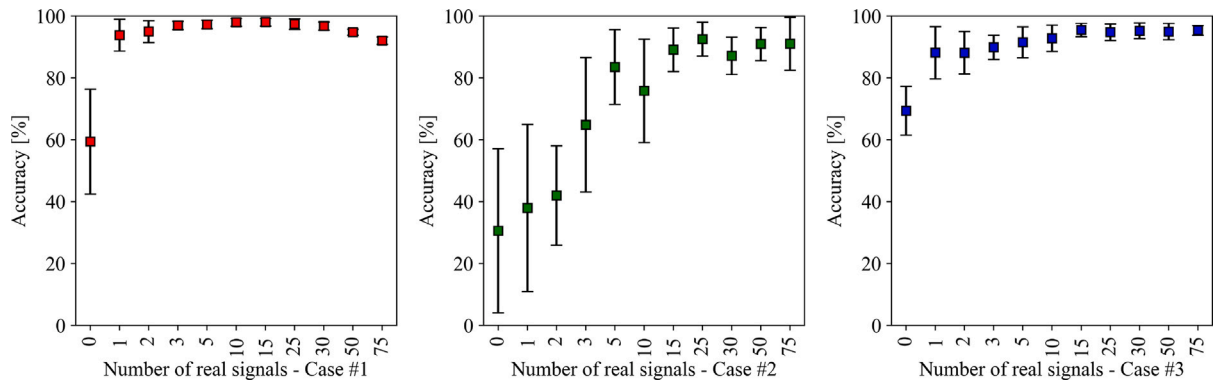

Fig. 7. Results for each case varying the amount of real signals used to create the training set.

a greater amount of signals in training for learning. Note that there was no significant gain in the accuracy and stability of the model (less standard deviation) in increasing the number of signals above 5250 signals, and therefore, the amount was selected for the other analyses.

5.3. Number of real samples for training

Based on the previous analysis, 5250 signals were used in the training, aiming to balance the computational cost and the efficiency of the model.

Unlike the previous analysis, the number of real signals to generate the total amount of training was varied. The goal is to analyze whether, through zero or few data acquisitions, it is possible to train a model robust enough for fault diagnosis.

11 conditions were tested, using 0, 1, 2, 3, 5, 10, 15, 25, 30, 50 and 75 real signals to generate the 5250 signals in the training. The values were selected so that the number of times the signals were augmented resulted in integer values, with no need to discard excess signals. As Case 2 has only 104 real signals per condition, the range was set from 1 to 75 signals. The results obtained for accuracy and standard deviation are shown in Fig. 7.

Unlike the other conditions, in the first test, all signals for the training group were generated synthetically, without performing the combination with the real signal (zero number of real signal). It can be seen in Fig. 7, that with zero real signal the lowest performing test was achieved. This result was expected, since the signals were created, basically containing the machine rotation frequency, white noise and random frequencies with amplitude close to the noise. This characteristic does not represent the dynamic behavior of the machine, resulting in the lowest value obtained. Therefore, it can be concluded that using the chosen strategy, it is not possible to train a model using only synthetic data without the combination with the real signal. It is worth mentioning that, despite not being the focus of the work, other

strategies, such as more complex modeling, can be tested to overcome this problem.

Analyzing Fig. 7, it can also be seen that using up to two real signals to create the entire training dataset, the average accuracy is significantly lower than the others, and the model has a large standard deviation. This occurs because if one of the selected signals presents some noise or excitation not related to the dynamic behavior of the machine, such pattern will be replicated for at least 50% of the training set signals, leading the model to error.

The increase in the number of signals tends to contribute to the improvement in the performance of the model, although it does not mean a linearly increasing behavior. It is noted that, in general, there is a stabilization of accuracy and standard deviation after 15 signals. In view of the computational and acquisition cost, and the results obtained with the increase in the amount of real data, it is suggested to be between 15 and 30 signals, a good working range for FaultD-XAI.

The maximum accuracy value reached in Case 1 was 98.1% and standard deviation of 1.2% with 15 signals. For Case 2, the maximum accuracy value was 92.5% and standard deviation of 5.5% with 25 signals. Case 3 obtained the highest accuracy value with 15 signals, being 95.5% and standard deviation of 2.1%. Smaller values of standard deviation were obtained for other amounts of signals, however none of them had a difference greater than 0.5% in relation to those presented.

The differences in accuracy and standard deviation obtained between each case are in accordance with what was expected in the study. Case 1 is a bearings remaining useful life test, where data were collected continuously, without the bench being dismantled to insert the defects. As there were no changes in the test bench, variations in the dynamic behavior of the machine are not expected, other than those caused by the wear of the components (since it is isolated). This makes the fault identification simpler for the model.

On the other hand, in Case 2 and 3, the benches were dismantled to insert the faults. Therefore, small changes in relation to the real

Table 1
Confusion matrix — Case 1.

Case 1	Normal	BPFO	BPFI	Unbalance	Misalignment	Looseness	Gear fault
Normal	52.01% (0.51%) 504 (5)	0.72% (0.41%) 7 (4)	0% (0%) 0 (0)	0% (0%) 0 (0)	0.41% (0.51%) 4 (5)	0.10% (0.10%) 1 (1)	0% (0%) 0 (0)
BPFO	0.62% (0.92%) 6 (9)	46.03% (0.92%) 446 (9)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0.10% (0.10%) 1 (1)	0% (0%) 0 (0)
BPFI	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)
Unbalance	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)
Misalignment	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)
Looseness	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)
Gear fault	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)

signal used as a reference for generating the synthetic data may occur during assembly and adjustment, generating a reduction in accuracy. This condition represents a situation where the system was trained with a machine signal in a similar condition, but not identical to the current one (which is a transfer learning problem definition).

The two analyzes are extremely interesting, and show the applicability of the system in the industrial scenario. In Case 1, it addresses the monitoring of a machine, where the system was installed, trained, and over time was able to identify a fault in the asset. In Case 2 and 3, the model was initially installed and trained. Hypothetically, the machine underwent some corrective maintenance, and the same model, without being retrained, continued to monitor the asset, and was able to identify the faults present. The fact that there is no need to retrain the model for small changes in the asset or reestablish the operating condition through corrective maintenance, overcomes a major problem in applications that is retraining. It is worth mentioning that for major interventions or changes in projects that modify the dynamic behavior of the machine, the model will tend to reduce its accuracy if not retraining.

The confusion matrix for the best values obtained in Case 1 is presented, containing the number of signals per class and the standard deviation for the 10 tests, in addition to the value in percentage, Table 1. Where in the column are the True Labels and in the row are the Predicted Labels.

In Table 1, it can be seen that the model confused normal signals with BPFO, misalignment and looseness. On the other hand, the faulted signals (BPFO) were confused with the normal condition and looseness only. The confusion between normal and BPFO may still be acceptable, due to the run to failure test feature and manual label. Therefore, signals referring to the beginning of the fault may present similar characteristics. The other classifications were model errors. Analyzing the result only by class, the normal signals were classified correctly 97.67% of the time, and faulty signals (BPFO) 98.45%. The confusion matrix for Case 2 is shown in Table 2. In Case 2, Table 2, normal condition and gear fault were confused with unbalance, misalignment, looseness and gear defect. As the gear fault characteristics are different from the errors presented, it can be concluded that there is a model error, despite the low amount. Analyzing by class, normal signals were classified correctly 91.14% of the time, and faulty signals 92.67%. The confusion matrix for Case 3 is shown in Table 3.

Table 3 presents the results obtained for Case 3. The normal signals were confused with all types of faults, despite the low amount. The unbalance signals were confused with normal condition, BPFO and looseness. Misalignment was confused with normal condition, unbalance, looseness and gear fault. Finally, looseness was confused only with unbalance and misalignment. Among the classes, the normal condition was the one with the highest accuracy rate with 99.04%.

The misalignment showed the highest error, resulting in 90.56%. The unbalance showed 98.97% and looseness 93.49%. The results show the greatest difficulty in classifying misalignment and looseness, which was expected, since some harmonics (such as: 2x and 3x) can be highlighted in both condition, making the analysis difficult.

5.4. Supervised training with real signals only

A comparative analysis was performed to verify the difference in accuracy and robustness of the model using FaultD-XAI and with supervised training using only real data.

In supervised training, all samples have labels and the signals used for each class refer to real fault conditions. Although in general, it presents excellent results, it is necessary to have samples of all fault conditions and they need to be previously labeled. Assumptions that often make industrial applications unfeasible. The real signals were divided into 70% training and 30% testing, and as in the proposed approach, they were executed 10 times due to the randomness of the model. All available signals were used. The accuracy and standard deviation for each test were: 98.1% (0.53%), 98.9% (0.86%) and 99.8% (0.11%), for Case 1, 2 and 3 respectively.

The result obtained for Case 1 it was the same using the two types of training (supervised with real signals and FaultD-XAI). The only difference was lower the standard deviation (0.53% compared to 1.2% in the proposed methodology). It is believed that the results were similar due to the lower complexity of the dataset, and consequently ease of identification of the fault. Moreover, the faults have more specific pattern compared to the other cases.

In Cases 2 and 3, supervised training with real data obtained higher results and lower standard deviation than FaultD-XAI, as expected. In Case 2 the precision was improved from 92.5% to 98.9% and the standard deviation decreased from 5.5% to 0.86%. In Case 3, the accuracy was 95.5% with FaultD-XAI becoming 99.8% using only the real signals, and the standard deviation went from 2.1% to 0.11%.

The results confirm the expected superiority of supervised training with only real data. On the other hand, it is worth mentioning that obtaining real and labeled data of all faults can be a complex and time-consuming task in the industrial applications, problems that are overcome by FaultD-XAI.

5.5. Explainable artificial intelligence (XAI)

Understanding how the AI model arrived at the end result is critical to its application in the industry. The models need a minimum of interpretability, so that users can trust in the predictions.

In addition to allowing us to understand the most relevant parameters used by the model for the final classification, the analysis also allows us to evaluate the learning of the model.

Table 2

Confusion matrix — Case 2.

Case 2	Normal	BPFO	BPFI	Unbalance	Misalignment	Looseness	Gear fault
Normal	7.90% (0.98%) 72 (9)	0% (0%) 0 (0)	0% (0%) 0 (0)	0.11% (0.11%) 1 (1)	0.11% (0.11%) 1 (1)	0.11% (0.11%) 1 (1)	0.44% (0.76%) 4 (7)
BPFO	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)
BPFI	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)
Unbalance	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)
Misalignment	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)
Looseness	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)
Gear fault	5.49% (3.29%) 50 (30)	0% (0%) 0 (0)	0% (0%) 0 (0)	0.88% (2.19%) 8 (20)	0.22% (0.32%) 2 (3)	0.11% (0.11%) 1 (1)	84.63% (7.68%) 771 (70)

Table 3

Confusion matrix — Case 3.

Case 3	Normal	BPFO	BPFI	Unbalance	Misalignment	Looseness	Gear fault
Normal	24.54% (0.17%) 1223 (9)	0.04% (0.09%) 2 (5)	0.02% (0.02%) 1 (1)	0.06% (0%) 3 (0)	0.04% (0%) 2 (0)	0.02% (0.02%) 1 (1)	0.06% (0.09%) 3 (5)
BPFO	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)
BPFI	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)
Unbalance	0.08% (0.04%) 4 (2)	0.02% (0.02%) 1 (1)	0% (0%) 0 (0)	24.82% (0.39%) 1247 (20)	0% (0%) 0 (0)	0.16% (0.39%) 8 (20)	0% (0%) 0 (0)
Misalignment	0.62% (0.65%) 31 (33)	0% (0%) 0 (0)	0% (0%) 0 (0)	0.02% (0%) 1 (0)	22.71% (2.46%) 1141 (124)	1.71% (2.40%) 86 (121)	0.02% (0%) 1 (0)
Looseness	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	1.53% (0.85%) 77 (43)	0.10% (0.13%) 5 (7)	23.44% (0.75%) 1178 (38)	0% (0%) 0 (0)
Gear fault	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)	0% (0%) 0 (0)

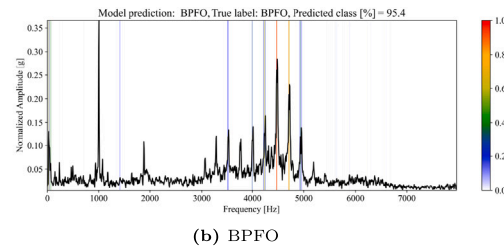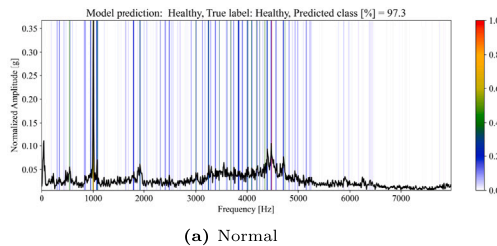**(a)** Normal **(b)** BPFO

Fig. 8. XAI analysis — Case 1.

The model's interpretability makes it possible to understand which were the most relevant features used for the prediction, and also to evaluate the model's learning.

The gradients used by the Grad-CAM method after the classification of the sample by the 1D CNN are evaluated. The heat map with the most relevant frequencies for the prediction are presented. A random sample was selected for each type of classification in each case, to exemplify the most relevant features, and validate if they are in accordance with the available literature on faults in vibration analysis of rotating machines. The results obtained for Case 1 are shown in Fig. 8.

Analyzing Fig. 8(a), it is verified that the normal signal uses several frequencies of the signal, not having a specific frequency or region. This was to be expected, as no faults are present, the entire signal referring to a normal machine condition. In the faulty condition, Fig. 8(b), the high frequency regions are excited, since it is a bearing fault. Consequently, the model uses the frequencies associated with the BPFO as the most relevant for indicating the defect. The analysis corresponds to a manual analysis performed by a rotating machinery specialist. The results obtained for Case 2 are shown in Fig. 9.

Case 2, Fig. 9(a), the model uses 1xGMF as a reference for determining the no-fault condition. It is normal to see peaks in GMF, even when the system does not present a fault, as identified by the method. On the other hand, when sidebands start to appear and harmonics appear in greater evidence, it is quite likely that a gear fault is happening. Note that in Fig. 9(b), the method identified the harmonics of the GMF and its sidebands as the most relevant frequencies, again validating the manual analysis of an expert.

The results obtained for Case 3 are shown in Fig. 10. The signals are zoomed in the 0 to 500 hz range for easy viewing. For the normal condition, Fig. 10(a), the model did not identify a specific frequency as expected. On the other hand, in Fig. 10(b), there is unbalance, where normally 1 x fr is dominant in the signal. Once again, the model presented the same analysis by an expert, indicating 1 x fr as the most relevant frequency. In Fig. 10(c), the most relevant frequency is 2 x fr, which is exactly one of the misalignment characteristics, where 2 x fr is greater than 1 x fr, and there may or may not be harmonics of 3x, 4x, 5x etc. The mechanical looseness can present itself in some ways, being one of them the excitation of the fr harmonics. In Fig. 10(d)
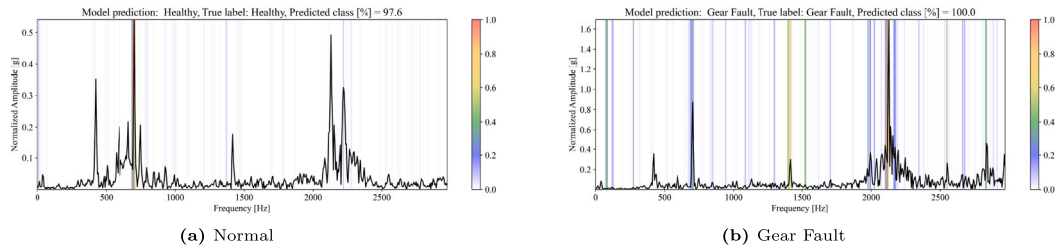
(a) Normal

(b) Gear Fault

Fig. 9. XAI analysis — Case 2.

(a) Normal

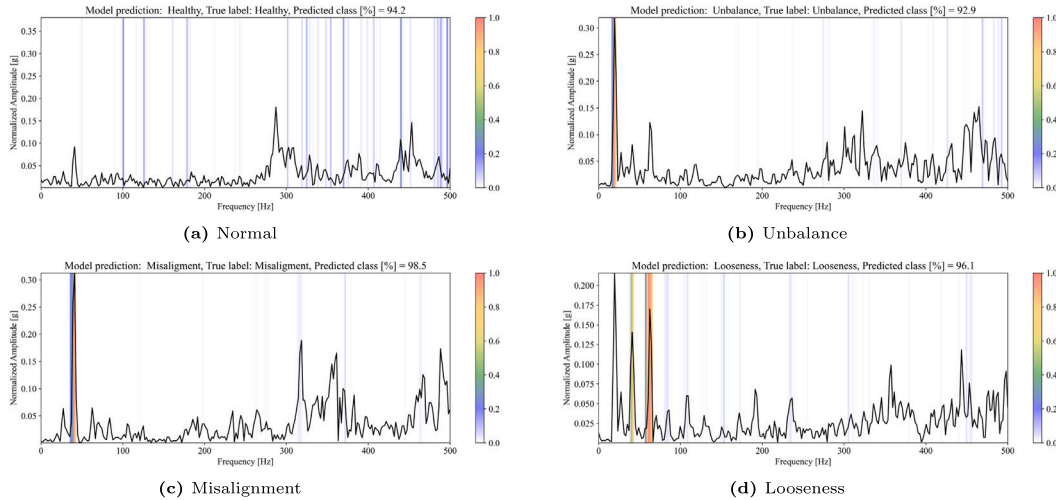**(b)** Unbalance

(c) Misalignment

(d) Looseness

Fig. 10. XAI analysis — Case 3.

it is noted that the model used the rotation harmonics as a reference, with emphasis on 3 x fr, again confirming the analysis based on human knowledge.

5.6. Limitations and future directions

The generation of synthetic faults depends on prior knowledge about the dynamic behavior of each fault, and although the vast majority are currently known, their modeling can be a limitation for the application of the method. The proposed method was based on the diagnosis of faults in rotating machines, with a focus on predictive maintenance. On the other hand, as suggested in Zhang et al. (2023), the use of more complex numerical models associated with the proposed method can help to obtain better results, optimizing the analysis of the dynamic behavior of equipment, as in Digital Twins applications.

Future work will also explore FaultD-XAI to remaining useful life estimation of rotating machinery components (Susto et al., 2014), using synthetic data, transfer learning and XAI. FaultD-XAI can also be exploited by modifying the models used to improve hit rates and computational complexity.

The use of Grad-CAM makes it possible to analyze the reasons why the model presents a classification error. Thus, future works will study the application of the method to improve the assertiveness of fault diagnosis methodologies.

6. Conclusions

This paper presents a new generic and interpretable approach to classifying faults in rotating machinery based on transfer learning from augmented synthetic data to real rotating machinery. The six-stage scheme is adopted: (i) Data Acquisition; (ii) Signal Generation; (iii) Data Augmentation; (iv) Signal Processing; (v) Fault Diagnosis;

(vi) Explainable Artificial Intelligence (XAI). From the acquired vibration signals, the synthetic faults are generated and augmented. Subsequently, they are transformed to the frequency domain, and used to train the AI model. After training, the model is able to classify real fault signals, without them having been used in training. Finally, explainability allows analyzing the predictions obtained.

To the best of our knowledge, the proposed FaultD-XAI is the first work that combined synthetic data with data augmentation to avoid the need of real data by exploiting XAI and transfer learning for fault diagnosis in rotating machinery. The main contributions of FaultD-XAI are: (i) a new transfer learning classification approach based on a synthetic dataset, without the need to have signals of real fault conditions; (ii) possibility of interpreting the way in which the final result was obtained by the model, supporting decision making (a new contribution to the study of XAI in fault diagnosis); (iii) a generic and simple way to generate synthetic fault data for training, based on the knowledge available in vibration analysis, without the need for complex models; (iv) possibility of generating varied training datasets of different sizes (data augmentation-targeting deep learning applications); (v) new dataset, publicly available, to study failures such as: unbalance, misalignment and looseness; (vi) possibility of application in different types of faults; (vii) faster deployment of the model in production; (viii) industrial application on real world datasets.

The results show that it is possible to train a model for fault diagnosis in rotating machinery, without having labeled data for all faults, overcoming this major problem in industrial applications. In addition, FaultD-XAI makes it possible to identify the most relevant features used by the model to perform the prediction, overcoming another major problem, which is to trust the result obtained by black-box models.

In summary, the best results (accuracy and standard deviation) reached in Case 1, 2 and 3 were: 98.1% (1.2%), 92.5% (5.5%) and 95.5% (2.1%), respectively. Using the traditional supervised training

method and real data, the results obtained were: 98.1% (0.53%), 98.9% (0.86%) and 99.8% (0.11%), for Case 1, 2 and 3 respectively. Therefore, comparing both results, it can be concluded that FaultD-XAI presents a promissory proposal for fault diagnosis. Moreover, the model's explainability analysis shows that it was able to learn the most relevant characteristics of each condition, and that FaultD-XAI is capable of supporting the expert's final decision-making.

As FaultD-XAI does not require labeled data of all possible machine operating conditions, and only knowledge currently available on fault diagnosis through vibration analysis, the methodology has many possible engineering applications.

CRediT authorship contribution statement

Lucas Costa Brito: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Analysis and interpretation of data, Writing – original draft, Writing – review & editing. **Gian Antonio Susto:** Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing. **Jorge Nei Brito:** Conceptualization, Writing – review & editing. **Marcus Antonio Viana Duarte:** Conceptualization, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

I have shared a link with the dataset.

Acknowledgments

References

Abdeljaber, O., Avci, O., Kiranyaz, M., Boashash, B., Sodano, H., & Inman, D. (2017). 1-D CNNs for structural damage detection: ?verification on a structural health monitoring benchmark data. *Neurocomputing, 52*.

Alves, D. S., Daniel, G. B., Castro, H. F., Machado, T. H., Cavalca, K. L., Gecgel, O., Dias, J. P., & Ekwaro-Osire, S. (2020). Uncertainty quantification in deep convolutional neural network diagnostics of journal bearings with ovalization fault. *Mechanism and Machine Theory, 149*, Article 103835.

An, Y., Zhang, K., Chai, Y., Liu, Q., & Huang, X. (2023). Domain adaptation network base on contrastive learning for bearings fault diagnosis under variable working conditions. *Expert Systems with Applications, 212*, Article 118802.

Appana, D., Ahmad, W., & Kim, J.-M. (2017). Speed invariant bearing fault characterization using convolutional. *Neural Networks, 40*, 189–198.

Asutkar, S., Chalke, C., Shivgan, K., & Tallur, S. (2023). Tinyml-enabled edge implementation of transfer learning framework for domain generalization in machine fault diagnosis. *Expert Systems with Applications, 213, Part B*, Article 119016.

Avci, O., Abdeljaber, O., Kiranyaz, S., Hussein, M., & Inman, D. (2018). Wireless and real-time structural damage detection: A novel decentralized method for wireless sensor networks. *Journal of Sound and Vibration*.

Brito, L., Susto, G., Brito, J., & Duarte, M. (2021). Fault detection of bearing: An unsupervised machine learning approach exploiting feature extraction and dimensionality reduction. *Informatics, 8*(4), 85.

Brito, L. C., Susto, G. A., Brito, J. N., & Duarte, M. A. V. (2022). An explainable artificial intelligence approach for unsupervised fault detection and diagnosis in rotating machinery. *Mechanical Systems and Signal Processing, 163*, Article 108105.

Cao, P., Zhang, S., & Tang, J. (2018). Preprocessing-free gear fault diagnosis using small datasets with deep convolutional neural network-based transfer learning. *IEEE Access, 6*, 26241–26253.

Chen, H., Hu, N., Cheng, Z., Zhang, L., & Zhang, Y. (2019). A deep convolutional neural network based fusion method of two-direction vibration signal data for health state identification of planetary gearboxes. *Measurement, 146*, 268–278.

Chuya-Sumba, J., Alonso-Valerdi, L., & Ibarra-Zarate, D. I. (2022). Deep-learning method based on 1D convolutional neural network for intelligent fault diagnosis of rotating machines. *Applied Sciences, 12*, 2158.

Ciabattoni, L., Ferracuti, F., Freddi, A., & Monteriù, A. (2018). Statistical spectral analysis for fault diagnosis of rotating machines. *IEEE Transactions on Industrial Electronics, 65*(5), 4301–4310.

Dalvand, F., Dalvand, S., Sharafi, F., & Pecht, M. (2017). Current noise cancellation for bearing fault diagnosis using time shifting. *IEEE Transactions on Industrial Electronics, 64*(10), 8138–8147.

Ding, X., & He, Q. (2017). Energy-fluctuated multiscale feature learning with deep ConvNet for intelligent spindle bearing fault diagnosis. *IEEE Transactions on Instrumentation and Measurement, 66*, 1926–1935.

Duan, L., Xie, M., Wang, J., & Bai, T. (2018). Deep learning enabled intelligent fault diagnosis: Overview and applications. *Journal of Intelligent & Fuzzy Systems, 35*, 5771–5784.

Feng, F., Wu, C., Zhu, J., Wu, S., Tian, Q., & Jiang, P. (2020). Research on multitask fault diagnosis and weight visualization of rotating machinery based on convolutional neural network. *Journal of the Brazilian Society of Mechanical Sciences and Engineering, 42*(11), 1–14.

Gecgel, O., Dias, J., Ekwaro-Osire, S., Alves, D. S., Machado, T. H., Daniel, G. B., & Cavalca, K. L. (2021). Simulation-driven deep learning approach for wear diagnostics in hydrodynamic journal bearings. *Journal of Tribology, 143*, 8.

Gentner, N., Carletti, M., Kyek, A., Susto, G. A., & Yang, Y. (2021). DBAM: Making virtual metrology/soft sensing with time series data scalable through deep learning. *Control Engineering Practice, 116*, Article 104914.

Guo, L., Lei, Y., Xing, S., Yan, T., & Li, N. (2019). Deep convolutional transfer learning network: A new method for intelligent fault diagnosis of machines with unlabeled data. *IEEE Transactions on Industrial Electronics, 66*(9), 7316–7325.

Handikherkar, V. C., & Phalle, V. M. (2021). Gear fault detection using machine learning techniques-A simulation-driven approach. *International Journal of Engineering, 34*, 212–223.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*.

Hoang, D., & Kang, H. (2019). A survey on deep learning based bearing fault diagnosis. *Neurocomputing, 335*, 327–335.

Ince, T., Kiranyaz, S., Eren, L., Askar, M., & Gabbouj, M. (2016). Real-time motor fault detection by 1-D convolutional neural networks. *IEEE Transactions on Industrial Electronics, 63*, 7067–7075.

ISO (2009). *Mechanical vibration — evaluation of machine vibration by measurements on non-rotating parts — Part 3* (ISO 10816-3:2009 ed.). International Organization for Standardization.

Jan, Z., Ahamed, F., Mayer, W., Patel, N., Grossmann, G., Stumptner, M., & Kuusk, A. (2023). Artificial intelligence for industry 4.0: Systematic review of applications, challenges, and opportunities. *Expert Systems with Applications, 216*, Article 119456.

Khan, A., Hwang, H., & Kim, H. S. (2021). Synthetic data augmentation and deep learning for the fault diagnosis of rotating machines. *Mathematics, 9*, 2336.

Khan, S., & Yairi, T. (2018). A review on the application of deep learning in system health management. *Mechanical Systems and Signal Processing, 107*, 241–265.

Kim, J., & Kim, J.-M. (2020). Bearing fault diagnosis using grad-CAM and acoustic emission signals. *Applied Sciences, 10*, 2050.

Kiranyaz, S., Avci, O., Abdeljaber, O., Ince, T., Gabbouj, M., & Inman, D. J. (2021). 1D convolutional neural networks and applications: A survey. *Mechanical Systems and Signal Processing, 151*, Article 107398.

Kiranyaz, S., Ince, T., Abdeljaber, O., Avci, O., & Gabbouj, M. (2019). 1-d convolutional neural networks for signal processing applications. In *ICASSP 2019-2019 IEEE international conference on acoustics, speech and signal processing* (pp. 8360–8364). IEEE.

Krizhevsky, A., Sutskever, I., & Hinton, G. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems, 25*, 1097–1105.

Kumar, P., & Hati, A. S. (2020). Review on machine learning algorithm based fault detection in induction motors. *Archives of Computational Methods in Engineering*, 1–12.

Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE, 86*(11), 2278–2324.

Lei, Y. (2017). *Intelligent fault diagnosis and remaining useful life prediction of rotating machinery*. Elsevier Inc.

Lei, Y., Yang, B., Jiang, X., Jia, F., Li, N., & Nandi, A. K. (2020). Applications of machine learning to machine fault diagnosis: A review and roadmap. *Mechanical Systems and Signal Processing, 138*, Article 106587.

Li, X., Zhang, W., & Ding, Q. (2019). Understanding and improving deep learning-based rolling bearing fault diagnosis with attention mechanism. *Signal Processing, 161*, 136–154.

Li, X., Zhang, W., Ding, Q., & Sun, J.-Q. (2020). Intelligent rotating machinery fault diagnosis based on deep learning using data augmentation. *Journal of Intelligent Manufacturing, 31*, 433–452.

Li, C., Zhang, S., Qin, Y., & Estupinan, E. (2020). A systematic review of deep transfer learning for machinery fault diagnosis. *Neurocomputing, 407*, 121–135.

Lin, C. J., & Jhang, J. Y. (2021). Bearing fault diagnosis using a grad-cam-based convolutional neuro-fuzzy network. *Mathematics, 9*(13), 1502.

Liu, R., Yang, B., Zio, E., & Chen, X. (2018). Artificial intelligence for fault diagnosis of rotating machinery: A review. *Mechanical Systems and Signal Processing, 108*, 33–47.

Lu, C., Wang, Z., & Zhou, B. (2017). Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification. *Advanced Engineering Informatics, 32*, 139–151.

Qiu, H., Lee, J., Lin, J., & Yu, G. (2006). Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics. *Journal of Sound and Vibration, 289*(4), 1066–1090.

Randall, R. B., & Antoni, J. (2011). Rolling element bearing diagnostics—A tutorial. *Mechanical Systems and Signal Processing, 25*, 485–520.

Samuel, P. D., & Pines, D. J. (2005). A review of vibration-based techniques for helicopter transmission diagnostics. *Journal of Sound and Vibration, 282*(1), 475–508.

Saufi, S. R., Ahmad, Z. A. B., Leong, M. S., & Lim, M. H. (2019). Challenges and opportunities of deep learning models for machinery fault detection and diagnosis: A review. *IEEE Access, 7*, 122644–122662.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. In *2017 IEEE international conference on computer vision* (pp. 618–626).

Shen, F., Chen, C., Yan, R., & Gao, R. (2015). Bearing fault diagnosis based on SVD feature extraction and transfer learning classification. In *Prognostics and system health management conference* (pp. 1–6). PHM.

Sobie, C., Freitas, C., & Nicolai, M. (2018). Simulation-driven machine learning: Bearing fault classification. *Mechanical Systems and Signal Processing, 99*, 403–419.

Susto, G. A., Schirru, A., Pampuri, S., McLoone, S., & Beghi, A. (2014). Machine learning for predictive maintenance: A multiple classifier approach. *IEEE Transactions on Industrial Informatics, 11*(3), 812–820.

Wang, X., Mao, D., & Li, X. (2021). Bearing fault diagnosis based on vibro-acoustic data fusion and 1D-CNN network. *Measurement, 173*, Article 108518.

Wu, X., Zhang, Y., Cheng, C., & Peng, Z. (2021). A hybrid classification autoencoder for semi-supervised fault diagnosis in rotating machinery. *Mechanical Systems and Signal Processing, 14*, Article 107327.

Xu, J., Zhou, L., Zhao, W., Fan, Y., Ding, X., & Yuan, X. (2022). Zero-shot learning for compound fault diagnosis of bearings. *Expert Systems with Applications, 190*, Article 116197.

Yang, B., Lei, Y., Jia, F., & Xing, S. (2019). An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings. *Mechanical Systems and Signal Processing, 122*, 692–706.

Yu, S., Wang, M., Pang, S., Song, L., & Qiao, S. (2022). Intelligent fault diagnosis and visual interpretability of rotating machinery based on residual neural network. *Measurement, 196*, Article 111228.

Zhang, J., Zhao, Y., Shone, F., Li, Z., Frangi, A. F., Xie, S. Q., & Zhang, Z. (2023). Physics-informed deep learning for musculoskeletal modeling: Predicting muscle forces and joint kinematics from surface EMG. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 31*.

Zhao, Z., Li, T., An, B., Wang, S., Ding, B., Yan, R., & Chen, X. (2022). Model-driven deep unrolling: Towards interpretable deep learning against noise attacks for intelligent fault diagnosis. *ISA Transactions, 129*, 644–662.

Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., & Gao, R. X. (2019). Deep learning and its applications to machine health monitoring. *Mechanical Systems and Signal Processing, 115*, 213–237.

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). Explainable machine learning in industry 4.0: Evaluating feature importance in anomaly detection to enable root cause analysis. In *Proceedings of the 2016 IEEE conference on computer vision and pattern recognition* (pp. 2921–2929). IEEE.

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., & He, Q. (2021). A comprehensive survey on transfer learning. *Proceedings of the IEEE, 109*, 43–76.

Lucas Costa Brito Consultant. PhD at UFU (Federal University of Uberlândia). Master's degree from UFU (Federal University of Uberlândia). Mechanical Engineer from UFSJ (Federal University of São João del-Rei). Green Belt at Six Sigma. Specialist in fault detection in industrial equipment through Predictive Techniques (Vibration Analysis, Thermographic Analysis, Oil Analysis, Grease Analysis, ODS — Operational Deflection Shape, Dynamic Balancing and Laser Alignment) and Artificial Intelligence (Artificial Neural Networks, Expert Systems and Hybrid Systems). Deputy Technical Director of FENEMI (National Federation of Mechanical and Industrial Engineering). Participant in the CsF (Science without Borders) Program at Lakehead University, Thunder Bay, Canada, from DEC 2013 to APR 2015. Member of LAV (Laboratory of Acoustics and Vibration) of UFU under coordination of Prof. D.Sc. Marcus Antonio Viana Duarte. Collaborator of Gep_LASID (Group of Studies and Research of the Laboratory of Dynamic Systems) of UFSJ under coordination of Prof. D.Sc. Jorge Nei Brito. Author of technical and scientific articles presented and published in national and international technical-scientific conferences. Guest speaker at regional and national conferences.