Development of a methodology for diagnosing faults in bearings operating under variable operating conditions based on self-supervised learning

Racquel Knust Domingues[1], Júlio A. Cordioli[2], and Danilo Silva[3]

[1,2,3] *Federal University of Santa Catarina, Florianópolis, Santa Catarina, 88040-900, Brazil*
racquel.knust@lva.ufsc.br
julio.cordioli@ufsc.br
danilo.silva@ufsc.br

1. Problem Contextualization

Predictive maintenance is crucial for ensuring the efficiency and availability of industrial assets by analyzing their current state to predict failures and enable timely corrective actions. Among the existing industrial assets, rotating elements are commonly used, leading to widespread utilization of bearings, as they are essential for reducing friction in rotary motion (Lei, 2016). Despite the many existing methods for detecting and diagnosing faults in these elements, the increasing complexity of systems due to technological advancements has led to a greater diversity of operating conditions, requiring new diagnostic methods.

Fault diagnosis methods can be classified into physical model-based, prior knowledge-based, and data-based methods (X. Zhang, Zhao, & Lin, 2021). The first two rely on a deep understanding of the element's physical behavior and can become complex and challenging to implement. In contrast, data-based methods have proven efficient by extracting useful maintenance information directly from the measured data of the asset's internal parameters.

Data-based methods consist of three main steps (Mushtaq, Islam, & Sohaib, 2021). The first step concerns data acquisition. In this stage, an internal parameter of the asset, vibration will be considered in thus study, must be measured and stored for later analysis. In the second step, the process of attribute extraction and selection occurs. Here, various techniques can be used, and the goal is to extract representations and metrics that make it possible to distinguish between data from different conditions. The final step involves analyzing the attributes obtained in the previous step, allowing for the determination of the asset's current condition and establishing a solid foundation for maintenance decision-making.

Traditionally, the steps of attribute extraction and selection, as well as state classification, were performed manually, relying on the expertise of analysts knowledgeable about the behavior of the machine element. With the development of machine learning algorithms, this process began to be automated, making fault diagnosis methods increasingly robust (S. Zhang, Zhang, Wang, & Habetler, 2020). Initially, traditional shallow learning algorithms, such as SVM, were applied only in the classification step, which still required manual extraction and selection of attributes from the data processing. Subsequently, with the development and use of deep learning methods, the attribute extraction process also became automatic, as these methods have the capability to automatically extract the most useful representations for the task to which they are applied.

Deep learning algorithms applied to fault diagnosis typically rely on supervised learning, which requires labeled data. The performance of these algorithms improves with the number of labeled samples (Long, Chen, Yang, Huang, & Li, 2023). However, labeling large datasets can be costly and impractical due to diverse operating conditions. To address these challenges, researchers are exploring self-supervised learning (Chowdhury, Rosenthal, Waring, & Umeton, 2021), which leverages unlabeled data to learn useful representations from artificial labels created from the data itself. These representations can then be transferred to specific fault diagnosis tasks, allowing the model to achieve high performance with fewer labeled samples.

2. Theoretical Background

To learn useful representations of the data within self-supervised learning, a pretext task must be defined, and deep learning models are used to solve it. Additionally, this methodology involves the automatic creation of labels for unlabeled data according to the defined pretext task (Morningstar et al., 2024). A pretext task can be defined as a task aimed at extracting the inherent patterns in the data, that is, extracting features that make the data unique or similar, regardless of the subsequent task they are applied to.

Pretext tasks can be defined using several methods, including context-based, contrastive learning, and generative algorithms (Gui et al., 2023). Context-based methods focus on designing tasks to identify or characterize specific transformations in the data, assuming that the ability to do so allows the model to extract useful representations. Contrastive learning aims to maximize the similarity between representations from different views of the same sample while distinguishing them from other data, helping the model to recognize intrinsic patterns. Generative algorithms, on the other hand, work by reconstructing the input from a modified version, enabling the model to learn robust representations that capture both local and global features.

Based on the output of the previous process, the learned model can then be transferred to solve the target fault diagnosis task. At this stage, a dataset with a significantly smaller amount of labeled data can be used to fine-tune the model, as it is assumed that the best representations have already been learned in the previous phase. This transfer of learning can be done in two ways: either by retraining only the part of the model responsible for producing the expected result for the target task or by retraining the entire model (Ericsson, Gouk, Loy, & Hospedales, 2022). It is important to emphasize that this retraining is done exclusively to adapt the model to the target task. With this fine-tuning, a semi-supervised fault diagnosis methodology employing self-supervised learning is defined. Figure 1 presents a scheme of this methodology.

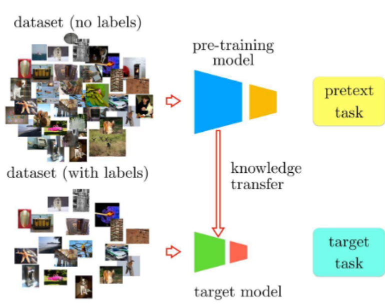

Figure 1. Schematic methodology employing self-supervised learning (Noroozi et al., 2018)

3. RELATED WORKS

In order to understand the current state of the literature on semi-supervised methodologies that use self-supervised learning, a systematic bibliographic review was carried out. From the studies identified, three main aspects were mapped: the format of the input data, the method used to define the pretext task, and the knowledge transfer process. It is worth mentioning that the search was focused exclusively on studies that used vibration signals to detect and diagnose bearing failures, that is, that had classification as the target task.

H. Wang, Liu, Ge, and Peng (2022) proposes a framework where the pretext task is defined as recognizing transformations of the raw input signal. This means that the self-learning phase involves applying certain signal processing techniques to the input data and training the model to recognize which technique was applied, fitting into a context-based method. The transfer learning is performed by partially freezing the model and retraining it for the target task using a smaller labeled dataset. It is noted that by defining the pretext task in this way, the model can extract generalized representations of the input data, as it needs to understand the intrinsic characteristics of the data to recognize its transformation.

Li, Wu, Deng, Wei, and Xu (2022) used self-supervised learning based on a contrastive learning method called Deep InfoMax (DIM). In this method, the objective of the pretext task is to maximize the mutual information between the input data and their local patches to obtain useful representations. Additionally, the article proposes the use of a signal-to-image conversion method, which generates grayscale images from raw signal segments to serve as input for the model. The retraining and fine-tuning are done using the model resulting from the self-learning process, which is frozen, along with a fully connected layer that is trained to perform the target task.

Yan and Liu (2022) proposed the use of a self-supervised architecture based on momentum contrastive learning (MoCo) for developing a fault diagnosis methodology for aircraft engine bearings. By utilizing contrastive learning for the pretext task, the goal is to identify similarity between different transformations of the same signal (positive pairs) or dissimilarity between transformations of different signals (negative pairs). To achieve this, the article proposes the use of signal multimodal learning, which employs augmentation techniques in two domains: time and frequency. The objective is to consider that the transformations of the same signal in these two domains are positive pairs, regardless of which domain they are represented in, and vice versa. For the knowledge transfer stage, the obtained model is frozen, and an SVM model is proposed as the classifier, thereby performing retraining and fine-tuning for the target task.

Wan, Chen, Zhou, and Shi (2022) also proposed a methodology using contrastive learning as the basis for self-learning, but using a Siamese network architecture and raw signals as input. In this case, the same input data is fed into two identical networks, with a different augmentation technique applied to each network. The results obtained from each network are compared, aiming to maximize the similarity between them. The retraining and fine-tuning stage in this methodology is done using a fully connected layer attached to the network trained in the previous stage, which is kept frozen.

W. Zhang, Chen, and Kong (2021) propose a methodology for fault diagnosis in bearings based on self-supervised learning using generative algorithms. The article suggests constructing an RGB input image using three signal-to-image transformations: wavelet transform, spectrogram, and time-domain signal-to-image conversion. The pretext task is defined as reconstructing a part of the input image that is omitted. This allows the model to learn intrinsic characteristics of the data. Additionally, the article proposes a joint learning framework, where supervised training is conducted simultaneously with self-supervised training, eliminating the need for the two-step procedure used in other methodologies.

4. MOTIVATION

It has been noted that the methodologies presented in the literature for fault diagnosis in bearings using self-supervised learning are predominantly applied and tested on data from constant operating conditions. However, in most cases, industrial assets operate with variable speeds and loads, which results in vibration signals that have non-stationary and often nonlinear characteristics (Z. Wang, Liu, Wu, & Wang, 2024).

These peculiarities of the signal can cause certain uncertainty and ambiguity, hindering the extraction of useful representations from the data that can effectively distinguish between a healthy condition and a faulty one. In such situations, a common approach is order analysis, which involves transforming the signal analysis domain from the time domain to the angular domain (Z. Wang, Yang, & Guo, 2022). This way, the analysis is conducted relative to a specific rotation, and even if there are fluctuations in rotational frequency, the information remains understandable because it is related to the rotation itself rather than an absolute value.

To apply order analysis, angular resampling can be used, which is a numerical technique that utilizes a reference signal containing the shaft phase to perform the domain transformation. Traditionally, this reference signal was obtained through an external device, such as a tachometer. However, in certain cases, this can be impractical and represent an additional cost. Therefore, using techniques to estimate the rotational speed of a machine based on its own vibration can offer a solution to this problem (Domingues, Cordioli, & Braga, 2023).

5. THESIS PROPOSAL

Taking into consideration what has been stated so far, this thesis proposes the development of a methodology for diagnosing faults in bearings operating under variable operating conditions, using vibration signals and based on self-supervised learning. The goal is to develop an end-to-end methodology, where a raw vibration signal is taken as input and the output is the diagnosis of the current condition of the bearing, both in terms of fault detection and in identifying its location and severity, when detected.

For the completion of this work, a division into five major stages was proposed, covering the necessary steps to achieve the expected final result. In Table 1, there is a schedule for carrying out these stages. The first of these concerns the systematic literature review on topics encompassing the central theme of this thesis, aiming to build a theoretical framework for the proposed proposition. A brief summary of what was mapped was presented in the previous sections of this document.

Table 1. Stage Schedule

Stage	Period
Literature Review	08/2023 - 07/2024
Acquisition of Experimental Data	08/2024 - 01/2025
Methodology Implementation	02/2025 - 01/2026
Assessment and Adjustments	02/2026 - 07/2026
Writing and Defense	08/2026 - 07/2027

The second stage of this thesis involves obtaining vibration data to be used for training the methodology to be developed. For this purpose, a dataset is being created from a test rig for simulating faults in rotating machinery, available at the home institution of this work, encompassing various operating conditions. Figure 2 shows an image of the test rig in question. This dataset includes variations in fault locations, covering faults in the outer race, inner race, and rolling elements of the bearing, as well as variations in severity, with three defect sizes. Additionally, different rotational conditions are being used, ranging from constant rotation to variable rotation conditions (sinusoidal variation, increasing ramp, and decreasing ramp). Four different test rig configurations are being used: 1) standard configuration, consisting of driving a shaft supported by two bearings through a motor, 2) introduction of point load in the standard configuration, using a heavy rotor, 3) transmission of load from an electromagnetic brake via a pulley system, and 4) transmission of load from an electromagnetic brake applied directly to one end of the shaft. Furthermore, the introduction of faults in other system elements, such as rotor imbalance and base misalignment, is planned to simulate conditions closer to those of a real machine.

The third stage involves the implementation of the methodology itself. This phase of the project will unfold in three steps. First, the primary fully supervised architectures found in the literature for fault detection and diagnosis in bearings will be applied to a small labeled portion of the dataset to establish baseline performance metrics, with the goal of improving these results using self-supervised methodologies. Next, preliminary implementations of self-supervised learning frameworks already existing in the literature will be carried out. The main objective of this step is to map the key limitations of these frameworks when applied to signals with varying operational conditions.

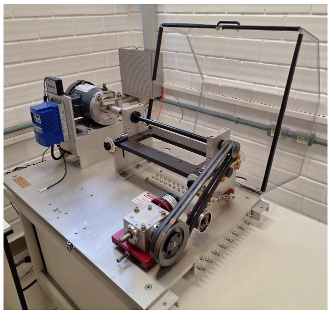

Figure 2. Simulated Fault Test Rig

Finally, a new methodology will be proposed and implemented. In this final part of the stage, two aspects will be investigated: the incorporation of order analysis concepts within the context of defining the pretext task and the use of different knowledge transfer techniques. Since pretext tasks using algorithms based on context prediction and contrastive learning are highly dependent on transformations, it is expected that the models can be further generalized, and their performance enhanced through the use of order analysis. Additionally, exploring different ways to transfer the knowledge obtained in the self-supervised learning stage may further contribute to improving the methodology's performance.

The fourth stage involves validating the methodology using data from the test rig. This includes training, validation, and necessary adjustments to optimize performance and generalization. The finalized methodology will then be validated on a real machine dataset provided by the project's partner company to measure its efficiency. Finally, the fifth stage includes writing and defending the thesis, along with the generation of scientific articles to present the research advancements.

REFERENCES

Chowdhury, A., Rosenthal, J., Waring, J., & Umeton, R. (2021). Applying self-supervised learning to medicine: review of the state of the art and medical implementations. In *Informatics* (Vol. 8, p. 59).

Domingues, R., Cordioli, J., & Braga, D. (2023). Evaluation of fault detection methodology in roller bearings using time-synchronous average. In *Proceedings of 27th abcm international congress of mechanical engineering*.

Ericsson, L., Gouk, H., Loy, C. C., & Hospedales, T. M. (2022). Self-supervised representation learning: Introduction, advances, and challenges. *IEEE Signal Processing Magazine*, *39*(3), 42–62.

Gui, J., Chen, T., Zhang, J., Cao, Q., Sun, Z., Luo, H., & Tao, D. (2023). *A survey on self-supervised learning: Algorithms, applications, and future trends.*

Lei, Y. (2016). *Intelligent fault diagnosis and remaining useful life prediction of rotating machinery*. Butterworth-Heinemann.

Li, G., Wu, J., Deng, C., Wei, M., & Xu, X. (2022). Self-supervised learning for intelligent fault diagnosis of rotating machinery with limited labeled data. *Applied Acoustics*, *191*, 108663.

Long, J., Chen, Y., Yang, Z., Huang, Y., & Li, C. (2023). A novel self-training semi-supervised deep learning approach for machinery fault diagnosis. *International Journal of Production Research*, *61*(23), 8238–8251.

Morningstar, W., Bijamov, A., Duvarney, C., Friedman, L., Kalibhat, N., Liu, L., ... others (2024). Augmentations vs algorithms: What works in self-supervised learning. *arXiv preprint arXiv:2403.05726*.

Mushtaq, S., Islam, M. M., & Sohaib, M. (2021). Deep learning aided data-driven fault diagnosis of rotatory machine: A comprehensive review. *Energies*, *14*(16), 5150.

Noroozi, M., Vinjimoor, A., Favaro, P., & Pirsiavash, H. (2018). Boosting self-supervised learning via knowledge transfer. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 9359–9367).

Wan, W., Chen, J., Zhou, Z., & Shi, Z. (2022). Self-supervised simple siamese framework for fault diagnosis of rotating machinery with unlabeled samples. *IEEE Transactions on Neural Networks and Learning Systems*.

Wang, H., Liu, Z., Ge, Y., & Peng, D. (2022). Self-supervised signal representation learning for machinery fault diagnosis under limited annotation data. *Knowledge-Based Systems*, *239*, 107978.

Wang, Z., Liu, T., Wu, X., & Wang, Y. (2024). A fault diagnosis method based on ntfes-fcct for variable working condition bearing signals. *IEEE Sensors Journal*.

Wang, Z., Yang, J., & Guo, Y. (2022). Unknown fault feature extraction of rolling bearings under variable speed conditions based on statistical complexity measures. *Mechanical systems and signal processing*, *172*, 108964.

Yan, Z., & Liu, H. (2022). Smoco: A powerful and efficient method based on self-supervised learning for fault diagnosis of aero-engine bearing under limited data. *Mathematics*, *10*(15), 2796.

Zhang, S., Zhang, S., Wang, B., & Habetler, T. G. (2020). Deep learning algorithms for bearing fault diagnostics—a comprehensive review. *IEEE Access*, *8*, 29857-29881. doi: 10.1109/ACCESS.2020.2972859

Zhang, W., Chen, D., & Kong, Y. (2021). Self-supervised joint learning fault diagnosis method based on three-channel vibration images. *Sensors*, *21*(14), 4774.

Zhang, X., Zhao, B., & Lin, Y. (2021). Machine learning based bearing fault diagnosis using the case western reserve university data: A review. *Ieee Access*, *9*, 155598–155608.