A Self-supervised Learning Approach for Anomaly Detection in Rotating Machinery

Fabrizio De Fabritiis[1,2], Konstantinos Gryllias[1,2,3]

[1] *Department of Mechanical Engineering, Division LMSD, KU Leuven, Celestijnenlaan 300, Box 2420, 3001 Leuven, Belgium*

[2] *Flanders Make@KU Leuven, Celestijnenlaan 300, Box 2420, 3001 Leuven, Belgium*

[3] *Leuven.AI – KU Leuven Institute for AI, B-3000, Leuven, Belgium*

ABSTRACT

Early fault detection in rotating machinery needs careful expert analysis of vibration data for monitoring a component state. Online methods that automatically set a threshold and raise an alarm when the vibration signature is anomalous are meant to efficiently manage key assets in a preventive maintenance plan.

In recent years a focus has raised on data driven methods in parallel with the increasing attention towards machine learning and, particularly, deep learning models. In this regard, for rotating equipment components, an important aspect relates to labelled data scarcity for supervised training. On the other hand, the advent of the Internet of Things allows to gather data from multiple assets with relevant information on the asset state itself. Self-supervised learning methods in deep learning application are currently tackling this problem. Investigating Self-learning approaches to integrate domain knowledge and learn relevant features from unlabeled data is therefore important for condition monitoring applications.

In this paper a methodology is proposed based on cycle consistency representation learning for training an embedder network on univariate unlabeled data. In order to learn a distance metric in the embedding space the original data are transformed to generate sequences of augmented inputs to enforce learnable pattern similarity in the augmented pairs. A differentiable cycle-consistency loss is chosen to maximize the numbers of augmented pairs in the learned embedding space that have minimum features distance. The pretext task in the described self-supervised setting aims to train a feature extractor for discriminating dissimilar samples in the embedding space by a distance metric and to provide a useful representation for down-stream tasks.

The paper analyzes the performance of the approach for anomaly detection in rotating machinery. The methodology is tested on vibration data provided by the Center for Intelligent Maintenance Systems (IMS), University of Cincinnati, considering different accelerated life test campaigns. The data were collected to monitor the fault development in bearings and the model shows how the learned embedding space discriminates effectively anomalous samples from normal ones in the degradation stages of the bearings.

1. INTRODUCTION

In rotating machinery, critical mechanical components are commonly monitored with vibration sensors directly mounted onto the machine. Incipient faults are detectable observing the variations in the vibration pattern (Henriquez, Alonso, Ferrer, & Travieso, 2013) and the context in which the machine is operating. While displacement sensors and velocity sensors are preferred for specific applications, piezoelectric accelerometers are widely adopted in most cases for being affordable, small and sensitive to a wide range of frequencies (Bogue, 2013).

Early fault detection in rotating machinery relies on model-driven (Jalayer, Orsenigo, & Vercellis, 2021) or data-driven (Liu & Gryllias, 2021) methods. A methodology for early fault detection is effective if it enables the identification of anomalies in e.g. vibration signals which are symptoms of the dynamic forces originated from the initial development of a defect in the monitored component. As an example, assessing the state of gears and bearings is critical for drivetrains in complex systems, such as wind turbines. A sudden failure can cause extensive downtime or catastrophic failure of an entire wind turbine (Yan, Dunnett, & Jackson, 2023).

The exponential growth of computing power in the last two decades has brought a major research interest towards data-driven methods in condition monitoring. In this regard, the availability of data plays an important role. During a machine lifetime a few measurements are collected under a faulty state of a component. The imbalance between healthy data and anomalous samples has motivated unsupervised or semi-supervised approaches in anomaly detection, in particular One-class Classifiers (Bi et al., 2024), (Liu & Gryllias, 2021). Among them, the following have been extensively applied for machine fault detection in literature: One-class Support Vector Machines (OCSVM), Deep Support Vector Data Description (Deep SVDD), Autoencoder (AE) and its variants, such as the Variational Autoencoder (VAE).

Prior knowledge on the geometry of the monitored component, the shaft rotational speed and the sampling rate are taken into account for reliable model-driven approaches to capture relevant information content on the dynamics of a machine sub-system from a vibration signal. The load and the context in which the machine operates can affect the measurements and the dynamics and therefore can be also included in a model depending on the targeted task. Deep learning architectures are able to extract low level features from complex vibration signals at a feature extraction stage. Nevertheless, features extracted in a pre-processing stage from vibration signals can facilitate the successful accomplishment of a condition monitoring task. In (Cui, Bangalore, & Tjernberg, 2018) the Wavelet Transform is applied for pre-processing the vibration data. In (Chen, Mauricio, Li, & Gryllias, 2020) the Cyclic Spectral Coherence map is used as input for a fault detection model in a deep learning framework.

To avoid data labelling, which can be expensive and time consuming with human-annotated labels, unsupervised learning has been proposed. Recent developments in unsupervised learning have led to a subset of approaches classified as self-supervised (Gui et al., 2023). Self-supervised learning aims to learn discriminative features on the data regardless of their labels. A deep learning model is trained on a pretext task for representation learning. The focus on the pretext task differentiate self-supervised learning from general unsupervised learning. The learned representation is then used in a downstream task where the inputs are the features which are the outputs of the pre-trained model. The remaining paper is organised as follows. In Section 2 the proposed methodology is introduced. Moreover in Section 3 the methodology is applied on a publicly available dataset and the results are analytically presented. The paper closes with a conclusion at Section 4.

2. PROPOSED METHODOLOGY

The proposed methodology is based on data augmentation and self-supervised learning for training a convolutional neural network (CNN) to map an input in an embedding space with a selected metric. The distance metric is intended to discriminate effectively dissimilar samples. Each embedding extracted by the pre-trained model from a test set is matched with the closest embedding from a reference set of healthy samples. The distances in the embedding space are considered in this methodology as anomaly scores.

2.1. Data pre-processing

The vibration signals collected from an accelerometer positioned on a bearing housing are processed to obtain spectrograms. The discrete short-time Fourier transform (STFT) is applied to the raw time signal x_n with window w

$$STFT\left\{x_n\right\}(h,k) = X(h,k) = \sum_{n=0}^{N-1} x_{n+h} w_n e^{-i2\pi\frac{kn}{N}}$$

leading to a time-frequency representation in the complex plane with phase and magnitude. In the proposed methodology, the magnitude of the STFT is retained as input of the CNN.

2.2. Data augmentation

Image transformations like translation, crop, flip, rotation, contrast, blur, and color distortion are used for augmentation as state-of-the-art techniques (Russell & Wang, 2023). In the case of a time frequency representation only a subset of those transformation is applicable. Frequency masking and time masking augmentation have been proposed in literature (Park et al., 2019) (Kim, Han, & Ko, 2021). The transformation considered on the STFT magnitude for data augmentation are:

- Random scaling
- Random frequency masking

The first transformation scale the input with maximum value between 1 and 0.1. The second transformation masks half of the frequencies similarly either to a band pass filter or to a band-stop filter as shown in Figure 1.

2.3. Deep learning model

Multi-layer neural networks and, in particular, CNNs have been applied extensively in representation learning and classification tasks. CNNs require significantly lower number of model parameters than fully connected (FC) networks, making them less prone to overfit. In contrast to FC networks, CNNs learn optimal kernels by incorporating the inherent spatial relationships within the data. For this reason, FC networks are not suitable for data with a grid-like topology such

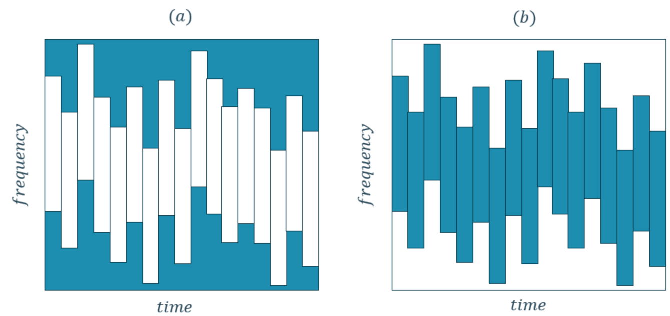
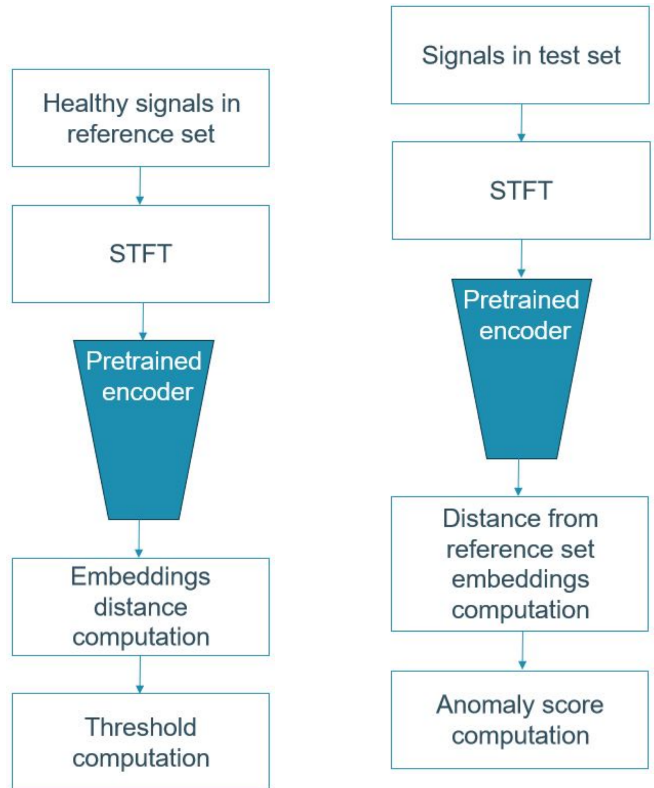


Figure 1. Frequency masking on a spectrogram. The masked frequencies are in white. In case (a) the frequencies are dropped out according to a band-stop filter. In case (b) the frequencies are dropped out according to a band pass filter. For each time bin the mask position is set randomly



(a) Threshold computation

(b) Anomaly score computation

Figure 2. (a) From a reference set the pre-trained encoder extract the embeddings and their distance is used to define an anomaly threshold. (b) In the test phase the anomaly score is computed as distance from the nearest neighbour in the reference set.

as images or time-series. With regard to condition monitoring, CNNs are preferred over FC networks for tracking the degradation of a system where sensors are placed to provide time series through a data acquisition system. Considering the topology of the spectrogram input of a convolutional neural network, with shape (h, l), the kernel size of the first layer of the network has been set according to Table 1.

Layers	Kernel shape	[Ch in,Ch out]	Activation
Convolutional	2D $(h \times 3)$	$[1, 10]$	ReLU
Convolutional	2D (1×3)	$[10, 20]$	ReLU
Convolutional	2D (1×3)	$[20, 40]$	ReLU
	Out features		
FC Linear	20	-	-

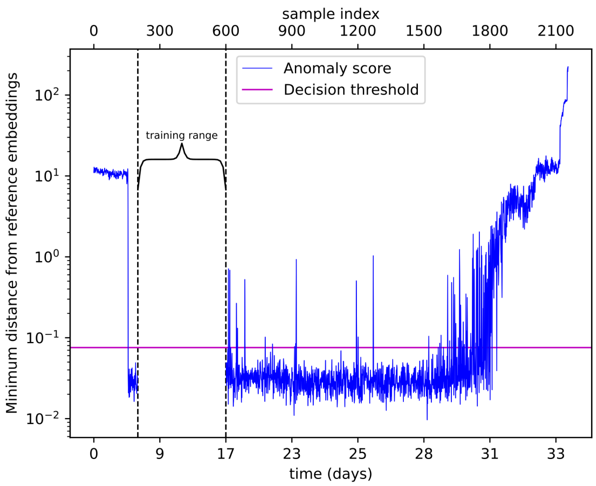
Table 1. CNN Hyperparameters for each layer

2.4. Cycle-consistency learning

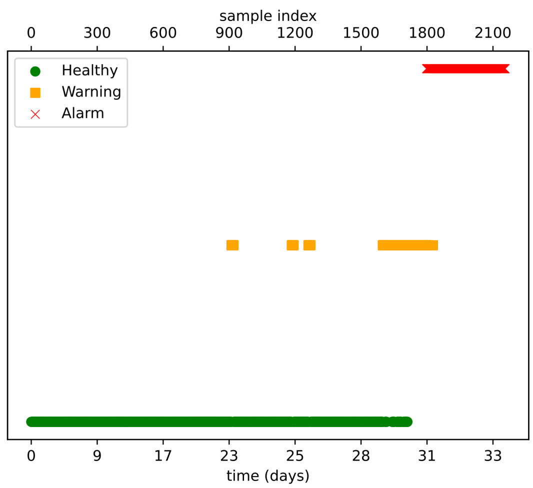
The proposed self-learning paradigm is based on extracting features that align two augmented views of the spectrograms in the training set. In order to map closely two transformed inputs from the same spectrogram, the number of points that can be paired from the two augmented views can be maximized by minimizing their distance in the learned embedding space (Dwibedi, Aytar, Tompson, Sermanet, & Zisserman, 2019). An embedding from one view is cycle-consistent when, once mapped to an embedding in the other view, is mapped back to itself. In other words, the embedding is cycle-consistent if it cycles back to itself. The embedding of a sample p_i in one augmented view $P = \{p_1, \ldots, p_N\}$ is the output of the neural network $f_i = \phi(p_i, \theta)$, being ϕ the neural network and θ its parameters. A point p_i is cycle-consistent if, considering another augmented view $Q = \{q_1, \ldots, q_N\}$ and the embeddings $G = \{g_1, \ldots, g_N\}$ where $g_i = \phi(q_i, \theta)$, the nearest neighbour $g_j = \arg min_{g \in G} \|f_i - g\|$ of its embedding f_i maps to the same embedding point when finding the nearest neighbour in $F = \{f_1, \ldots, f_N\}$ following $f_j = \arg min_{f \in F} \|g_j - f\|$. In order to train a neural network maximizing the cycle-consistent points between two views of N elements in the training set, the soft nearest neighbours $\tilde{f}, \tilde{g}$ are defined for a differentiable cycle consistency loss. The nearest neighbour of f_i is defined as $\tilde{g} = \sum_j^N \alpha_j g_j$ with $\alpha_j = \text{softmax}(-\|f_i - g_j\|)$. Considering a classification problem where each element of the view is a class itself, the cross entropy loss is minimized with predicted labels $\hat{y}_k = \text{softmax}(-\|\tilde{g} - f_k\|)$. The training procedure is described in Figure 11.

2.5. Anomaly detection

The pre-trained feature extractor maps the spectrograms in the learned embedding space. The distance metric enforced during training in the computation of the soft nearest neighbour is a dissimilarity measure used to discriminate embeddings for the task of anomaly detection. From a reference set of signals collected during a normal state of the machine component, the features of the spectrogram from the processed signal can be computed and a suitable threshold can be defined. For each spectrogram s_i in the reference set

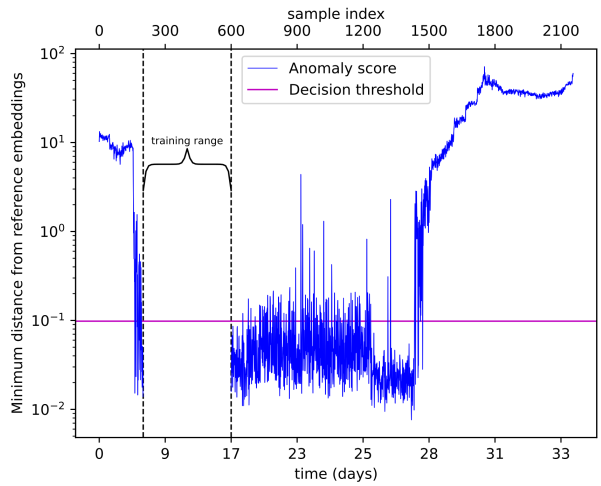


(a)

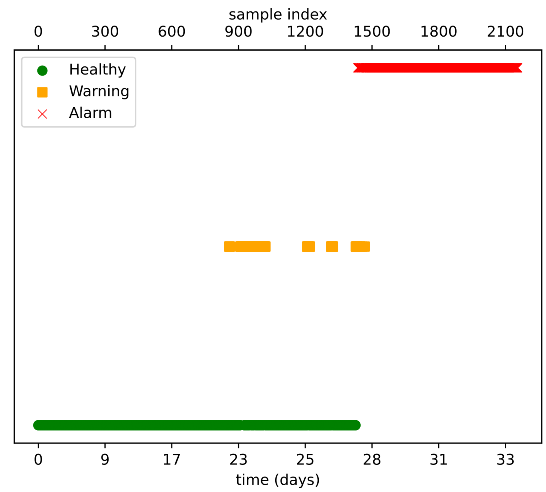


(b)

Figure 3. (a) Sensor level detection results for bearing 3 of dataset 1 and (b) bearing state recognition.



(a)



(b)

Figure 4. (a) Sensor level detection results for bearing 4 of dataset 1 and (b) bearing state recognition.

$S = \{s_1, \ldots, s_N\}$ the nearest neighbour is computed as

$$\hat{s}_i = \arg\, min_{s \in S \setminus \{s_i\}} \|\phi(s_i, \theta) - \phi(s, \theta)\| \tag{1}$$

where ϕ is the neural network with pre-trained parameters θ. Indicating the distance in the embedding space between s_i and the nearest neighbour s_i^{nn} in the reference set as

$$d_i = \|\phi(s_i, \theta) - \phi(\hat{s}_i, \theta)\| \tag{2}$$

The mean and variance of the distances of paired neighbours in the embedding space is

$$\mu = \frac{1}{N} \sum_{i}^{N} d_i \tag{3}$$

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{i}^{N} (d_i - \mu)^2} \tag{4}$$

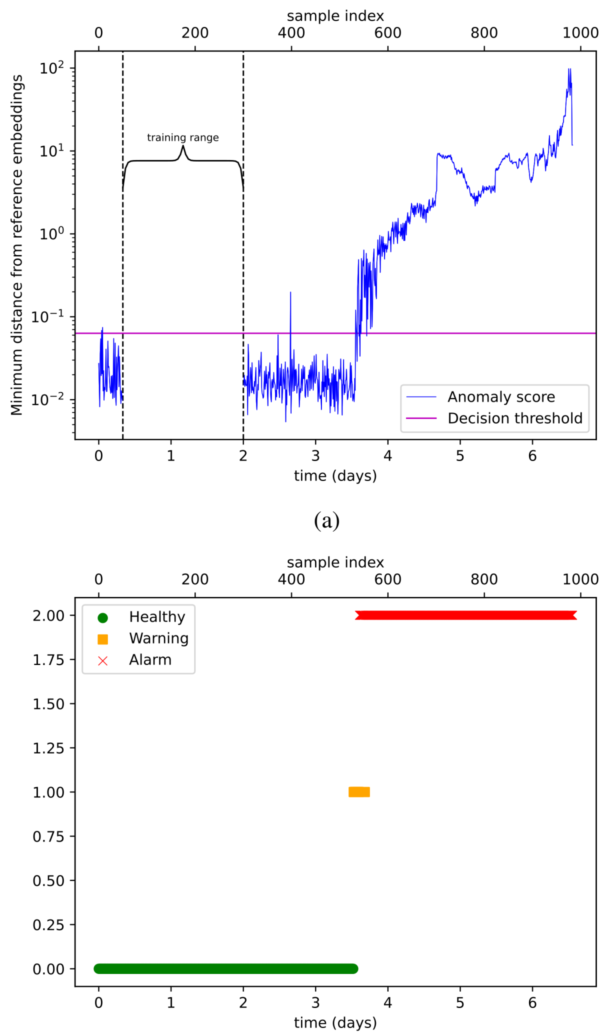
From the mean and the standard deviation a suitable threshold is defined as

$$d_{lim}^t = \mu + 3\sigma \tag{5}$$

as shown in Figure 2a. In order to perform early fault detection, the pre-trained model extracts the embeddings from the spectrograms in the test set. The distance from the nearest neighbour in the reference set embeddings

$$d_j^t = \|\phi(r_j, \theta) - \phi(\hat{s}_j, \theta)\| \tag{6}$$

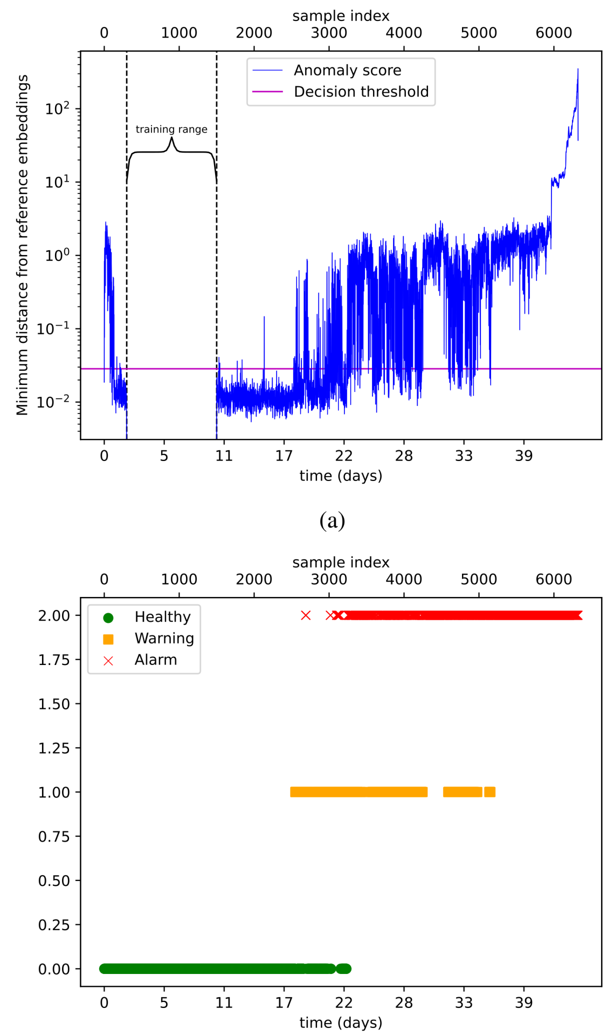
where $\hat{s}_j = \arg\, min_{s \in S} \|\phi(r_j, \theta) - \phi(s, \theta)\|$ and $r_j \in R$ is a spectrogram in the test set, is, in the proposed methodology, the anomaly score for a test sample according to Figure 2b.



(a)

(b)

Figure 5. (a) Sensor level detection results for bearing 1 of dataset 2 and (b) bearing state recognition.



(a)

(b)

Figure 6. (a) Sensor level detection results for bearing 3 of dataset 3 and (b) bearing state recognition.

3. APPLICATION OF THE METHODOLOGY

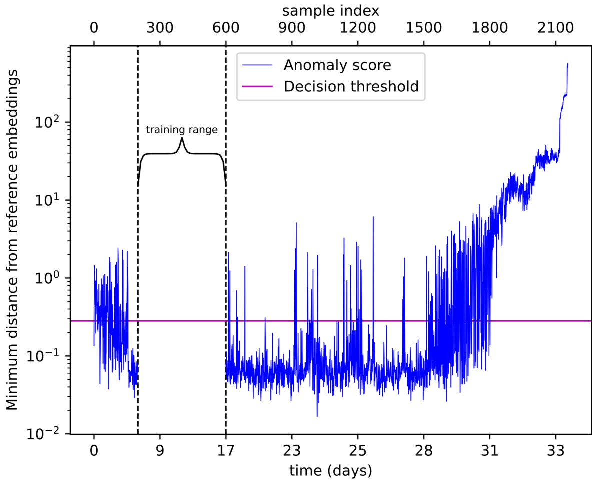
3.1. Dataset description

The methodology is applied and tested to the publicly available IMS bearing dataset, firstly introduced in 2006 (Qiu, Lee, Lin, & Yu, 2006) and provided by the Center for Intelligent Maintenance Systems, University of Cincinnati.

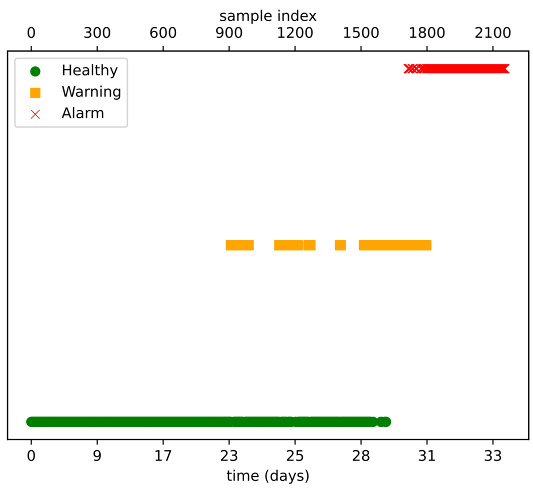
The dataset includes acceleration measurements from four separate accelerometers mounted on the housings of four Rexnord ZA-2115 double-row bearings installed on a single shaft as shown in Figure 12. The shaft was driven by an AC motor and maintained at a constant speed of 2000 RPM with a radial load of 6000 lbs with a belt transmission. The measurements have been collected with a sampling frequency of 20 KHz and each record refer to a 1-second vibration signal. Each record contains 20480 data points, indicating an actual sampling frequency of 20480 Hz. The measurements have been acquired at intervals of 5 minutes or 10 minutes. Some intervals between acquisitions are longer due to interruptions, therefore the recorded signals can be regarded as a discontinuous tracking of the system degradation.

The accelerated life tests were stopped when the accumulation of debris on a magnetic plug exceeded a certain level. Three run-to-failure tests have been conducted. At the end of each test all the bearings were examined to assess if and where faults occurred (Qiu et al., 2006):

- In the first run-to-failure test an inner race defect occurred in bearing 3 and a roller element defect in bearing 4.

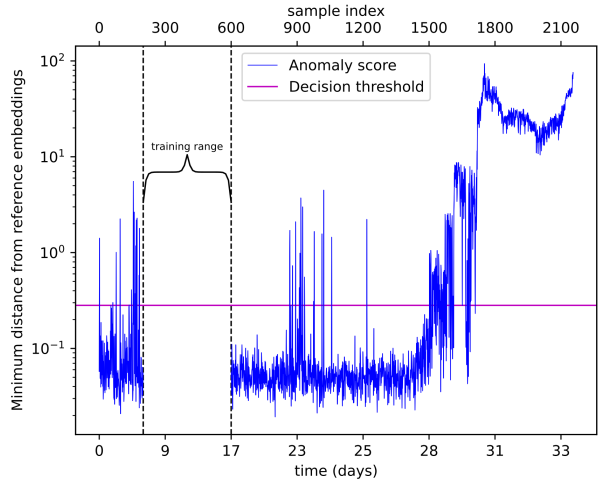


(a)

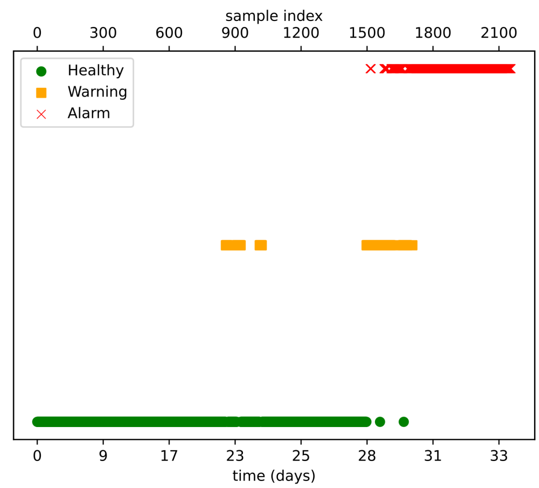


(b)

Figure 7. (a) Machine level detection results for bearing 3 of dataset 1 and (b) bearing state recognition.



(a)



(b)

Figure 8. (a) Machine level detection results for bearing 4 of dataset 1 and (b) bearing state recognition.

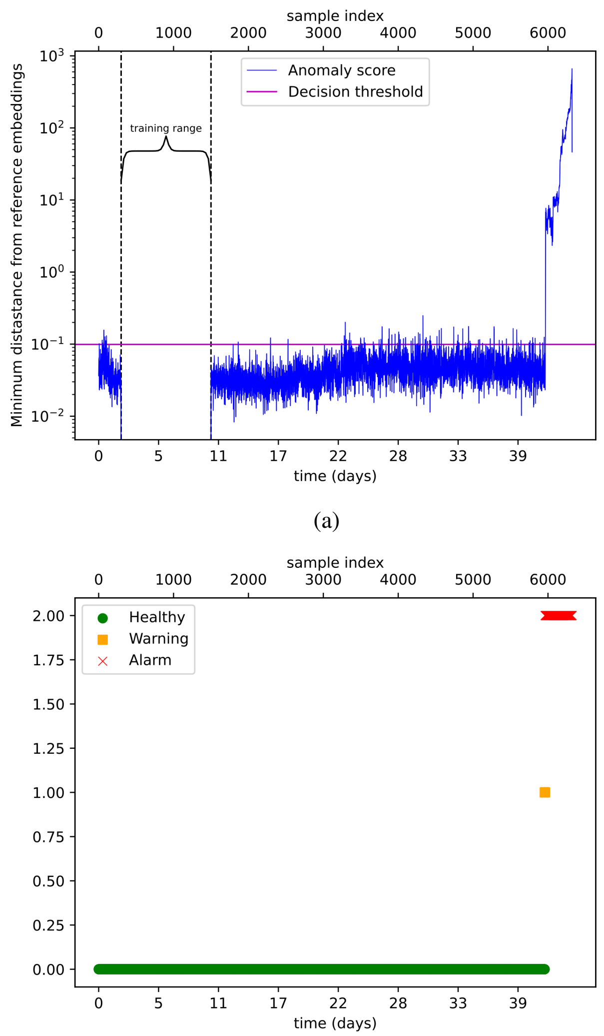
- In the second run-to-failure test an outer race defect occurred in bearing 1.
- In the third run-to-failure test an outer race defect occurred in bearing 3.

3.2. Input pre-processing

The window length of the STFT selected to process the 20480 data points in each recording is set to 1024 with an overlap on the previous segment of 50%. A Hamming window is applied for the Fast Fourier Transform (FFT) with chosen length of 1028 points. The output of the STFT is a time-frequency map of size $[515 \times 41]$. In the self-supervised training phase a random scaling and a random frequency masking is repeatedly applied on each input to obtain two different augmented views per epoch.

3.3. Detection strategy

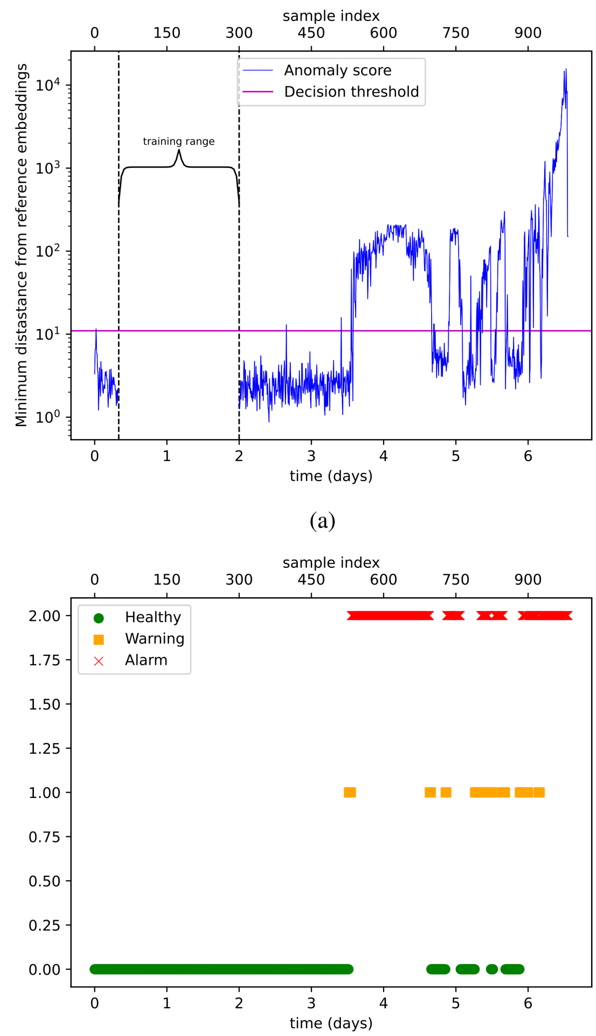
The distance metric chosen for the cycle consistency loss optimization is the euclidean distance. Once the encoder, with layers described in Table 3, is trained, the distances between the nearest neighbours in the embedding space of the training set are computed with Equation 2. The threshold is set according to Equation 5 with μ and σ calculated using the Equations 3 and 4 respectively. The anomaly scores of the test set are computed using the Equation 6 where the training set is the reference set of healthy samples. The fault detection strategy adopted is based on (Liu & Gryllias, 2020) where three conditions are proposed. Considering a moving window on the time history of the anomaly scores, the conditions are:



(a)

(b)

Figure 9. (a) Machine level detection results for bearing 3 of dataset 3 and (b) bearing state recognition.



(a)

(b)

Figure 10. (a) Machine level detection results for bearing 1 of dataset 2 and (b) bearing state recognition.

- At least 50% of the anomaly scores in the window are above the threshold.
- The average value of the anomaly scores in the window is above the threshold.
- At least 50% of the anomaly scores in the window are continuously above the threshold.

If all three conditions are satisfied the monitoring unit/component is considered in faulty state. If one or two conditions are satisfied, the state is considered a warning. The unit/component is considered in healthy state if none of the conditions is satisfied. The moving window length in this study consider the last ten anomaly scores. In an industrial application, a number of sensors can be mounted on a unit and a number of similar units are operating under similar operating conditions. The unit can be a factory machine, a wind turbine, a compressor etc. Therefore in the frames of condition monitoring and anomaly detection a question which arises is at which level should the training data be used: a) at a sensor level, where data are used to train a model specifically for this sensor or b) at a machine level, where data from all the sensors mounted on the unit are used to train a model at the unit level. In this paper, the effectiveness of the proposed anomaly detection model is evaluated at the two levels, thus the training datasets are prepared based on (a) data from a single sensor (sensor level) and (b) data from multiple sensors of one machine (machine level).

3.4. Sensor level early fault detection

All the training sets on which each model is trained under the self-supervised setting are supposed to contain samples

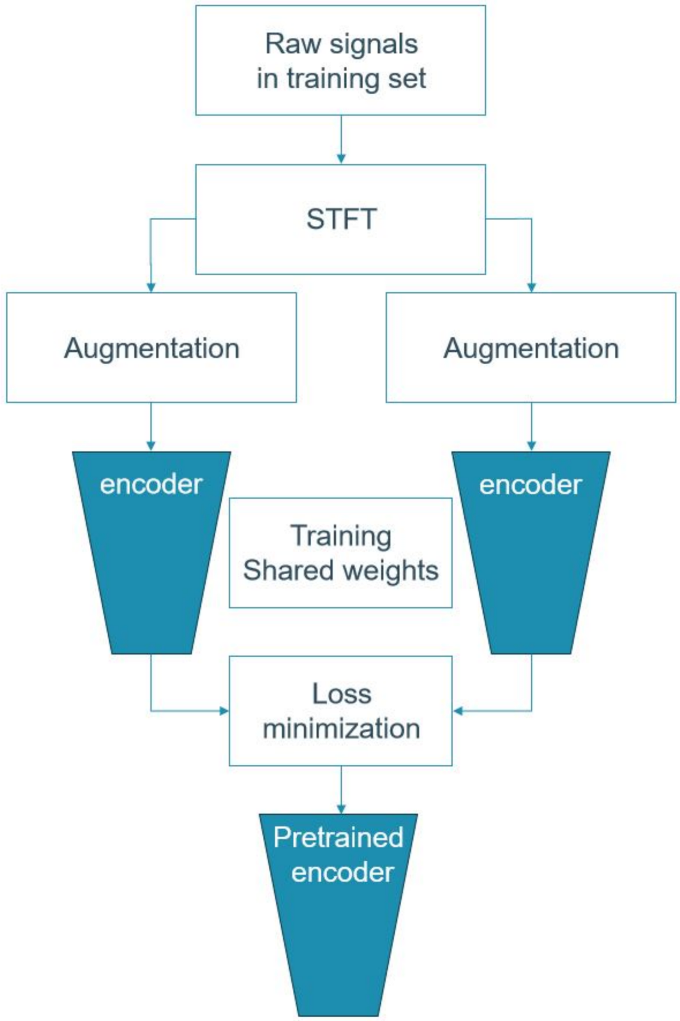


Figure 11. Structure of the self-supervised learning approach

acquired during the healthy state of the bearings. The training ranges for each dataset are set according to (Liu & Gryllias, 2020). In particular, the training set from the dataset contains the signals recorded from 7.45 days (record #200) to 17.05 days (record #600) for each sensor. With regard to this dataset, two accelerometers in orthogonal directions are placed on each housing. In this study only the data from one of the two accelerometers for each housing are considered, i.e. the data of the even numbered channels. For the second dataset the training set is composed of records collected from 0.35 days (#50) to 2.08 days (#300), while the training dataset of the third experimental campaign includes records acquired from 2.08 days (#300) to 10.41 days (#1500) by the start of the experiment. Each model is trained and tested on data acquired from only one sensor per dataset. The results of the detection on the faulty bearings are shown in Figures 3, 4 for dataset 1, Figure 5 for dataset 2 and Figure 6 for dataset 3. The first detection time of the faulty state is compared in Table 2 with the Support Vector Data Description method with negative samples (NSVDD) results reported in (Liu & Gryllias, 2020).

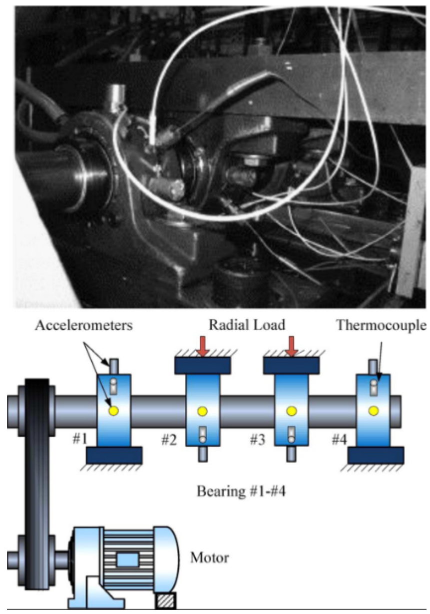


Figure 12. Schematic of the IMS bearing dataset test rig

	Proposed	(Liu & Gryllias, 2020)
Dataset 1 Bearing 3	**31.06 days** (#1808)	31.16 (#1823)
Dataset 1 Bearing 4	28.05 days (#1446)	**19.08 days** (#872)
Dataset 2 Bearing 1	3.83 days (#551)	**3.69 days** (#532)
Dataset 3 Bearing 3	**19.2 days** (#2697)	42.25 days (#5973)

Table 2. Sensor level detection comparison between the proposed model and the NSVDD model

3.5. Machine level early fault detection

In contrast to sensor level fault detection, the machine level fault detection aims to train a model from the acceleration responses of all the bearings of the machine. The results of the detection on the faulty bearings are shown in Figures 7, 8 for dataset 1, Figure 10 for dataset 2 and Figure 9 for dataset 3. The first detection time of the faulty state is compared in Table 2 with the results from the NSVDD results (Liu & Gryllias, 2020).

4. CONCLUSION

In this study, a new methodology for early fault detection based on vibration signals has been proposed. A feature extractor is trained in a self-supervised learning setting on unlabelled data. A distance metric is enforced with a cycle consistency loss optimization during training. The metric discrimi-

	Proposed	(Liu & Gryllias, 2020)
Dataset 1 Bearing 3	**30.47 days** (#1724)	31.26 (#1837)
Dataset 1 Bearing 4	29.08 days (#1524)	**28.33 days** (#1487)
Dataset 2 Bearing 1	**3.77 days** (#543)	5.11 days (#736)
Dataset 3 Bearing 3	**42.24 days** (#5972)	44.34 days (#6275)

Table 3. Machine level detection between the proposed model and the NSVDD model

nates faulty samples from healthy samples in the embedding space for the task of anomaly detection. The methodology is applied and evaluated on the publicly available IMS bearing dataset for early fault detection on data from one sensor or multiple sensors achieving high performance. The results show that the methodology can be trained using only a limited amount of training data and that it allows an easy comparison between vibration signals measuring (dis)similarity.

ACKNOWLEDGMENT

The authors gratefully acknowledge the European Commission for its support of the Marie Sklodowska Curie program through the H2020 ETN MOIRA project (GA 955681)

REFERENCES

Bi, Z., Yang, Y., Du, M., Yu, X., He, Q., & Peng, Z. (2024). Unsupervised hypersphere description approach for detecting and localizing anomalies in drivetrain with normal data. *Measurement*, 114349.

Bogue, R. (2013). Sensors for condition monitoring: A review of technologies and applications. *Sensor Review*, *33*(4), 295–299.

Chen, Z., Mauricio, A., Li, W., & Gryllias, K. (2020). A deep learning method for bearing fault diagnosis based on cyclic spectral coherence and convolutional neural networks. *Mechanical Systems and Signal Processing*, *140*, 106683.

Cui, Y., Bangalore, P., & Tjernberg, L. B. (2018). An anomaly detection approach using wavelet transform and artificial neural networks for condition monitoring of wind turbines' gearboxes. In *2018 power systems computation conference (pscc)* (pp. 1–7).

Dwibedi, D., Aytar, Y., Tompson, J., Sermanet, P., & Zisserman, A. (2019). Temporal cycle-consistency learning. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition* (pp. 1801–1810).

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. *arXiv preprint arXiv:2301.05712*.

Henriquez, P., Alonso, J. B., Ferrer, M. A., & Travieso, C. M. (2013). Review of automatic fault diagnosis systems using audio and vibration signals. *IEEE Transactions on systems, man, and cybernetics: Systems*, *44*(5), 642–652.

Jalayer, M., Orsenigo, C., & Vercellis, C. (2021). Fault detection and diagnosis for rotating machinery: A model based on convolutional lstm, fast fourier and continuous wavelet transforms. *Computers in Industry*, *125*, 103378.

Kim, G., Han, D. K., & Ko, H. (2021). Specmix: A mixed sample data augmentation method for training withtime-frequency domain features. *arXiv preprint arXiv:2108.03020*.

Liu, C., & Gryllias, K. (2020, 01). A semi-supervised support vector data description-based fault detection method for rolling element bearings based on cyclic spectral analysis. *Mechanical Systems and Signal Processing*, *140*.

Liu, C., & Gryllias, K. (2021). A deep support vector data description method for anomaly detection in helicopters. In *Phm society european conference* (Vol. 6, pp. 9–9).

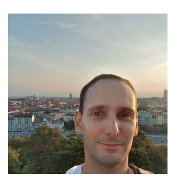
Park, D. S., Chan, W., Zhang, Y., Chiu, C.-C., Zoph, B., Cubuk, E. D., & Le, Q. V. (2019). Specaugment: A simple data augmentation method for automatic speech recognition. *arXiv preprint arXiv:1904.08779*.

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*(4-5), 1066–1090.

Russell, M., & Wang, P. (2023). Maximizing model generalization for machine condition monitoring with self-supervised learning and federated learning. *Journal of Manufacturing Systems*, *71*, 274–285.

Yan, R., Dunnett, S., & Jackson, L. (2023). Impact of condition monitoring on the maintenance and economic viability of offshore wind turbines. *Reliability Engineering & System Safety*, *238*, 109475.

BIOGRAPHIES



Fabrizio De Fabritiis received the B.S. degree and M.S. degree in mechanical engineering from Sapienza, University of Rome, in 2012 and 2017, respectively. He joined the LMSD Research Group, Department of Mechanical Engineering, KU Leuven, Leuven, Belgium, as a Ph.D. Researcher in 2022. His research interests include data-driven condition monitoring and machine learning application.

Konstantinos Gryllias received a 5 years Diploma and a Ph.D. degree in Mechanical Engineering from the National Technical University of Athens, Athens, Greece, in 2004 and 2010, respectively. He currently hols an Associate Professor position on Vibro-Acoustics of Machines and Transportation Systems at the Department of Mechanical Engineering, KU Leuven, Leuven, Belgium. He also serves as the Manager of the University Corelab Motion Products - MPro of Flanders Make, Belgium. His research interests include condition monitoring, signal processing, prognostics, and health management of mechanical and mechatronic systems.