Noise-aware AI methods for robust acoustic monitoring of bearings in industrial machines

Kerem Eryılmaz[1], Fernando de la Hucha Arce[2], Jeroen Zegers[3], and Ted Ooijevaar[4]

[1, 3, 4] *Flanders Make, Gaston Geenslaan 8, 3001 Leuven, Belgium*
kerem.eryilmaz@flandersmake.be
ted.ooijevaar@flandersmake.be

[2] *Flanders Make, Graaf Karel De Goedelaan 16-18, 8500 Kortrijk, Belgium*
fernando.delahuchaarce@flandersmake.be

ABSTRACT

Traditionally, companies have relied on vibration based condition monitoring technologies to implement condition based maintenance strategies. However, these technologies have drawbacks, such as the requirement of contact accelerometers. As an alternative, acoustic condition monitoring is non-invasive and allows for easy deployment. Furthermore, the use of microphones potentially enables the monitoring of multiple components using a single sensor, making the monitoring system scale better with machine or production complexity. However, microphone signals typically show a low signal-to-noise ratio (SNR), impacted by the high level of background noise which is often present in industrial environments. Particularly, the traditional method for monitoring the health condition of rolling element bearings, based on assessing whether the squared envelope spectrum of the bearing signal exceeds a given threshold at the fault frequencies, cause too many false positives when applied directly to microphone signals. It is therefore crucial to develop strategies to increase the robustness of acoustic monitoring methods. In this paper, we present and evaluate two data-driven strategies to robustly diagnose bearing faults from a microphone signal. Our proposed strategies are noise weighting based on the detection of background noise, and an artificial intelligence (AI) model that uses as input a combination of the traditional bearing fault frequencies and the mel spectrum of the microphone signal. These methods leverage both domain knowledge and data-driven techniques to increase the detection robustness. Our approach is implemented as a model trained and tested on bearing accelerated lifetime tests performed in the Smart Maintenance Lab setup at Flanders Make. Our results show that the use of our proposed strategies leads to significant improvements in diagnostic performance and time to first detection over noise-unaware acoustic monitoring methods.

1. INTRODUCTION

Condition monitoring involves the continuous monitoring of machine parameters to detect changes that are indicative of a developing fault. This a key component of condition based maintenance, the strategy that schedules maintenance actions based on the current health diagnosis of machine components, with the goal of reducing equipment downtime and total maintenance cost. The detection of faults in rolling element bearings is of special interest, since they are critical components of rotating machinery, and their faulty signals are often masked under other dominant sources (Randall, 2011). The use of accelerometers is the most common approach for monitoring bearing and gear faults, as vibrations often carry early information of their incipient damages (Lee et al., 2014).

There exist a wide range of well-established signal processing methods that are applied to vibration signals in order to estimate the health condition of a bearing. One of the most successful methods is envelope analysis, whose comprehensive description is given in (Randall & Antoni, 2011). It relies on the extraction and tracking of the fault characteristic frequencies in the squared envelope spectrum (SES) of the vibration signal generated by the bearing. As their name suggests, these frequencies are related to the bearing faults, and contain an increasing amount of energy as faults become more serious. For bearings operating under conditions of low load and low rotational speed, a different method based on stochastic resonance is proposed and shown to outperform envelope analysis in (Ompusunggu, Devos, & Petre, 2013).

However, a disadvantage of diagnosis techniques based on vibration analysis is that the accelerometers should be mounted close to the rotating component of interest. Consequently, several accelerometers are needed to monitor multiple bear-

ings, and accessibility constraints may render this impossible.

As an alternative to overcome these limitations, acoustic non-contact sensors such as microphones have recently drawn attention mainly for two reasons. First, they allow for easier deployment, as they do not need to be physically mounted by bolts, glue or magnets. Second, microphone signals may acquire information from several bearing signatures, potentially enabling the monitoring of multiple components with fewer sensors than in schemes based on accelerometers. Nevertheless, microphones will unavoidably collect signals from undesirable noise sources mixed with the signals emitted by the bearings. In industrial environments, these noise sources are generally quite strong and varied in nature, leading to microphone signals with low signal-to-noise ratio (SNR) that result in poor diagnosis performance. For this reason, dedicated methods to increase the robustness to background noise are crucial in acoustic monitoring.

Due to the wide array of different noise sources present in industrial environments, data-driven strategies are a powerful tool to increase the robustness of acoustic monitoring. In a data-driven strategy, healthy and damaged bearings are classified by a data-driven model trained using a set of relevant acoustics features. In (Mian, Choudhary, & Fatima, 2022), six sound quality features from microphone signals were used to train a support vector machine to diagnose bearing damages. For the diagnosis of bearing, rotor and stator faults in induction motors, a frequency domain feature extractor method combined with a nearest neighbour classifier is proposed and shown to perform well in (Glowacz, 2019).

Another commonly used strategy to achieve robust acoustic monitoring relies on microphone arrays and beamforming. The works presented in (Cardenas Cabada, Leclere, Antoni, & Hamzaoui, 2017; Ricardo Mauricio, Denayer, & Gryllias, 2022, 2023), and references therein, show that this strategy can produce good diagnosis results for bearing monitoring using beamforming. However, the requirement of multiple microphones and precise positioning increases the practical complexity of implementing this solution. For this reason, we consider beamforming strategies outside of the scope of this work, and focus on data-driven strategies using a single microphone.

In this paper we propose two data-driven methods to increase the robustness of the diagnosis of bearing faults using acoustic sensing. Our first approach is noise weighting based on the detection of background noise, and the second one is an artificial intelligence (AI) model whose input is a combination of the bearing fault frequencies and the mel-spectrum of the microphone signal. These methods integrate both domain knowledge and a data-driven technique, and they are trained and tested on bearing accelerated lifetime experiments performed in the Smart Maintenance Lab setup at Flanders Make. The goal is to evaluate the performance of our proposed methods, and show that acoustic monitoring is a cost effective and practical alternative to vibration monitoring.

The rest of this paper is structured as follows. In Section 2, the well-established envelope analysis method for bearing fault diagnosis is reviewed, and an explanation of its poor performance when applied to acoustic sensing is provided. Our two proposed data-driven methods for robust acoustic bearing fault diagnosis are detailed in Section 3. A description of the experimental setup is given in Section 4, which includes the performed bearing accelerated lifetime experiments, the acoustic scene, and the parameters of the signal processing and AI models. The performance of our proposed methods is evaluated and discussed in Section 5. Finally, the main conclusions are summarized in Section 6.

2. PROBLEM STATEMENT

Rolling element bearings are a crucial component in a wide variety of rotating machinery. However, over time they can develop faults such as surface fatigue defects or wear. For localized faults, as the rolling elements strike a fault in the inner or outer race, an impulse is generated that excites high frequency resonances on the structure between the bearing and the sensor location.

2.1. Vibration-based bearing fault diagnosis

The vibration signals from a faulty bearing can be modelled as a modulated blend of several signal components: an impulsive signal associated with the fault, the high frequency signals related to the the dynamics of other machine components such as the shaft and gears, the modulation between these signals and additional noise. The well-established method for bearing diagnostics is the so-called envelope method, which first enhances the impulsive signal generated by the fault, and then estimates the energy at the fault characteristic frequency and its harmonics from its squared envelope spectrum (SES). A complete explanation of the method is provided in (Randall & Antoni, 2011).

In this paper we focus on inner race faults, for which the fault characteristic frequency is the ball pass frequency, inner race (BPFI), given by

$$f_{\text{BPFI}} = \frac{n f_r}{2} \left\{ 1 + \frac{d}{D} \cos \phi \right\}, \tag{1}$$

where f_r is the shaft speed (frequency), n is the number of rolling elements, d is the diameter of the rolling elements, D is the pitch diameter , and ϕ is the contact angle. Other fault characteristic frequencies are the BPFO (ball pass frequency, outer race) and the BSF (ball spin frequency), corresponding respectively to outer race and rolling element faults.

In order to quantify the presence and severity of an inner race fault, we use as a feature the median of the SES value at the

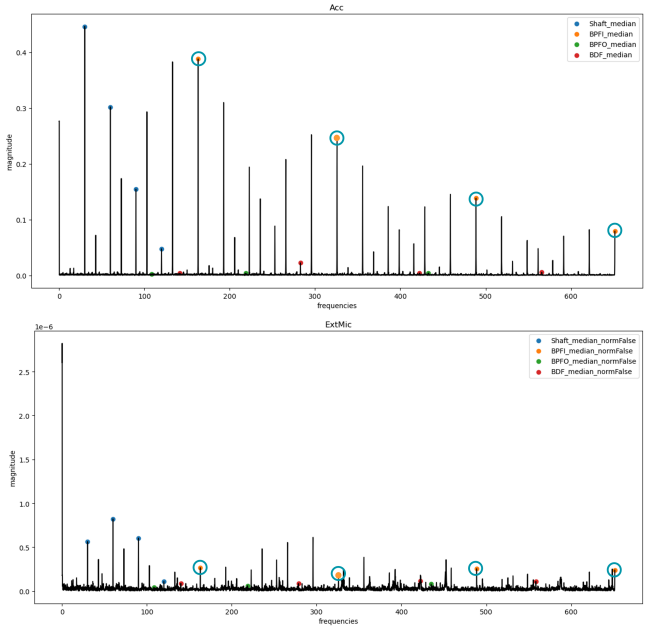

Figure 1. The squared envelope spectrum (SES) of the signal produced by a bearing with an inner race fault, acquired with an accelerometer (top) and a microphone (bottom). The BPFI and its first three harmonics are encircled in blue. The shaft frequency, BPFO and BDF (twice the BSF) are indicated for the sake of completeness.

BPFI and its harmonics. Throughout the rest of the paper, we refer to this feature as the BPFI feature. Mathematically, it is expressed as

$$\xi_{\text{BPFI}} = \underset{k}{\text{median}} \left\{ Y(k f_{\text{BPFI}}) \right\}, \quad k \in \{1, \ldots, n_{\text{harm}}\}, \quad (2)$$

where $Y(k f_{\text{BPFI}})$ denotes the peak magnitude of the SES at the k-th harmonic of the BPFI, $\underset{k}{\text{median}}\{\cdot\}$ denotes the median value of the set indexed by the integer k, and n_{harm} is the number of harmonics considered. Finding the peaks is done by searching the maximum SES magnitude around the theoretical fault frequency (Eq. 1) and its harmonics, within a pre-defined range tolerance.

2.2. Noise-unaware acoustic diagnosis

The direct application of diagnosis based on the fault characteristic frequencies, such as the BPFI, to acoustic sensing presents two problems. The first is that microphone signals are generally weaker than vibration signals, due to the larger distance between the microphone and the bearing. An example of this issue is provided in Figure 1, which shows a comparison of the SES from an accelerometer and a microphone signal produced by a bearing with an inner race fault. The BPFI and its first three harmonics can be easily identified in the accelerometer SES, while they cannot be clearly distinguished from the noise floor in the microphone SES.

Figure 2. Comparison of the BPFI feature ξ_{BPFI} between an accelerometer (top) and a microphone signal (bottom) over a bearing accelerated lifetime.

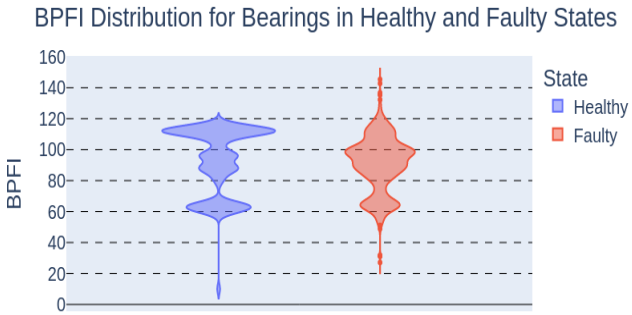

Figure 3. Comparison of the distribution of microphone-based BPFI feature ξ_{BPFI} between healthy and faulty states over the entire dataset.

The second problem is that the background acoustic noise is considerably stronger than the noise present in a vibration signal acquired by an accelerometer, and more diverse in nature due to the wide variety of potential noise sources present in industrial environments. As a result, microphone signals typically have a significantly poorer SNR. Moreover, due to this background noise, there will be additional energy present in the BPFI and its harmonics even when the bearing is healthy, leading to a great number of false positives over the bearing's lifetime.

This matter is illustrated in Figure 2, where a comparison is shown between the BPFI feature of an accelerometer and a microphone signal over the lifetime of a bearing in one of our accelerated lifetime experiments. The experimental setup and conditions are described in Section 4. It can be readily seen that the BPFI feature in the accelerometer signal displays a clear distinction between the healthy and faulty states of the bearing, while the BPFI feature in the microphone signal exhibits many spikes during the healthy state, leading to an unreliable diagnosis of the bearing inner race fault. Figure 3 further demonstrates the difficulty by showing the great overlap between the distributions of BPFI feature values acquired through the microphone for bearings in healthy and faulty states.

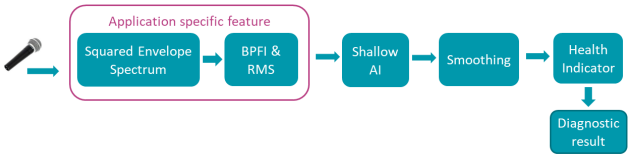

Figure 4. Diagram of the noise-unaware method with shallow AI for diagnosis of bearing inner race faults with acoustic sensing. The diagram would represent the baseline method by removing the shallow AI and the RMS feature.

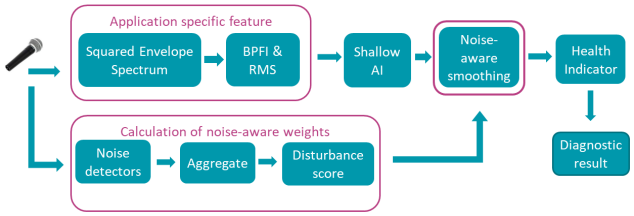

Figure 5. Diagram of noise-aware smoothing for diagnosis of bearing inner race faults with acoustic sensing.

The simplest strategy to enhance the diagnostic performance using the BPFI feature is to introduce a smoothing step. This method, with the choice of a median filter for smoothing, is what we consider our baseline for comparing the performance of diagnostic methods in this paper.

A more refined step is to introduce a shallow AI model before smoothing. In our case, this shallow AI is a two-layer fully connected neural network (NN) whose input features are the BPFI feature ξ_{BPFI} and the RMS value of the microphone signal. Both methods do not take the presence of acoustic noise explicitly into account, so we refer to them as noise-unaware methods. In particular, the BPFI and RMS features are very poor informants on the presence of background noise, hence the shallow AI can learn very little about rejecting undesired disturbances.

Figure 4 displays a diagram representing noise-unaware diagnosis with shallow AI and smoothing. The same diagram would represent the baseline method by removing the RMS feature and the shallow AI block.

3. Methods for robust acoustic bearing fault diagnosis

In this section we describe the two data-driven methods that we propose to increase robustness to noise in acoustic bearing fault diagnosis. As explained in Section 2, background noise introduces unreliability in the form of a high amount of false positives. The goal becomes therefore to reduce these false positives while retaining as much of the true positives as possible.

3.1. Noise-aware smoothing

Noise-aware smoothing aims to refine the health indicator calculated from the BPFI feature by taking into account the noise level present at each interval of time. To achieve this, a weighted median filter is applied to the raw health indicator over an interval of the last N points, where the weights are designed such that the influence of each point is inversely proportional to its noise level. A diagram of the diagnosis process including this strategy is shown in Figure 5.

The weighted median filter works as follows. If we are given a series of predictions $x_0, ..., x_t$ with noise levels $d_0, ..., d_t \in [0, 1]$, and a window size N, the noise-weighted prediction at time t would be computed as follows:

1. Assign a weight to every prediction $x_{t'}$ as $1 - d_{t'}$.
2. Sort predictions $x_{t-N}, ..., x_t$ and keep their associated weights $w_{t-N}, ..., w_t$ in that order too.
3. Compute the cumulative weight for each item $x_{t'}$, i.e. $\sum_{a=t-N}^{t'} w_a$.
4. The item where the cumulative weight exceeds half of the total weight is the weighted prediction, i.e. x'_t at time t such that $\sum_{a=t-N}^{t'} w_a \geq \frac{1}{2} \sum_{w=t-N}^{t} w_a$

Median smoothing follows the same procedure, except that all the weights are set to 1, reducing it to a regular median filter.

In order to obtain a noise level, each time interval is assigned a score that represents the likelihood that an undesired acoustic disturbance is present in it. For our case, an undesired disturbance is defined as any short-time sound that is not informative about the phenomenon being monitored, i.e., all sound events not generated by the bearing of interest itself. This excludes stationary background noise as well as disturbances that take last longer than a round of data acquisition (ten (10) seconds in our case).

This score is computed as the maximum of the output of a collection of noise detectors. These detectors are designed to indicate acoustic disturbances that can be characterized as events. This means any sounds whose presence in time, although it may be repeated, is limited. Examples include tools getting dropped, sporadic speech, various machinery turning on or off. Specifically, we implemented detectors targeting disturbances with the following characteristics:

Narrow-band disturbances: This detector indicates the presence of noise in a specific frequency band. In our experiments, the sources of this kind of disturbances were a pump and a forklift present in our laboratory.

High frequency disturbances: This detector indicates the presence loud, complex noises that have a lot of energy in high frequency bands. In our experiments, the source of this kind of disturbances was a neighboring experimental setup that kept loudly dropping off metal pipes.

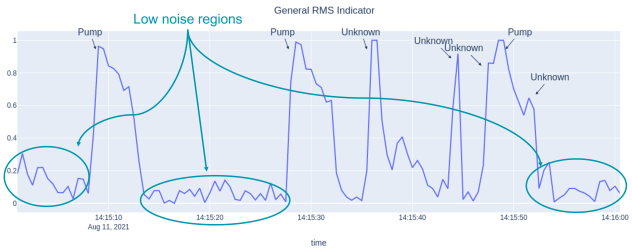

Figure 6. Example of the output of the detector for general loud disturbances. The regions circled correspond to intervals where no disturbance is detected.

General loud disturbances: A detector for loud events, designed to capture sudden changes in the root mean squared (RMS) value of the signal. This takes into account the fact that the bearings do not cause such changes at any point in their operation.

Speech disturbances: Voice Activity Detection (VAD) is a field of active research with many mature results. For this reason, we chose to utilize a VAD solution, provided by the Silero project (SileroTeam, 2021), based on a pre-trained neural network.

All detectors, except those for speech and general loud disturbances, require a characterization of the acoustic disturbances in the environment where the bearing of interest is located. An example of the output of the detector for general loud disturbances is shown in Figure 6. This detector captures some loud events that do not correspond to our known disturbance sources, marked as unknown. It also reacts to the pump activation, since it produces sudden changes in the RMS value of the signal. Capturing the same disturbance with several detectors is beneficial, as we are interested in catching as many as possible rather than determining their type. The regions of low disturbance score, that appear circled in the graph, are those given higher weight by noise-aware smoothing.

3.2. Noise awareness with deep AI and hybrid features

This method aims to utilize a deep AI model to obtain a reliable health indicator of bearing faults from acoustic information. The main idea is to introduce and train a deep AI model that uses adequately general features extracted from microphone recordings of healthy and faulty bearings. At a high level, it operates as a generalized way to clue the model in about what parts of the frequency spectrum are useful to pay attention to, and which parts are best to ignore. This model acts in combination with the very specific fault frequency features, thus integrating a data-driven technique with domain knowledge. The diagram in Figure 7 represents the diagnostic process that combines the indicators from both the BPFI feature and the deep AI-model.

The advantages of using a deep AI model are twofold. The first one is that it can learn complex patterns during the training process, leading to better diagnostic performance. The

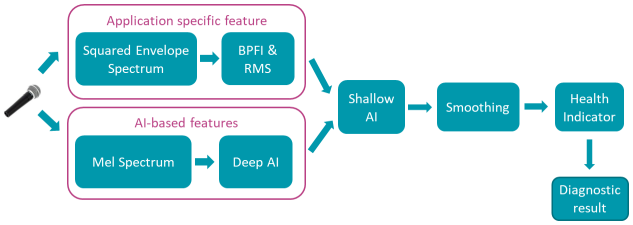

Figure 7. Diagram of deep AI-based noise awareness for diagnosis of bearing inner race faults with acoustic sensing.

second one is that, as long as sufficiently varied examples of disturbances are included during the training phase, it can learn to work with many kinds of noise sources.

The features for this method need to represent the relevant bearing fault information while being of reasonable dimensionality. For this purpose, the features we chose are the mel-spectrogram of the acoustic signal. This is a spectrogram obtained by a mel filter bank, a set of half-overlapped triangular filters equally spaced on the mel scale (Rabiner & Schafer, 2010). Since this is a logarithmic scale for frequency, the filters are narrower for low frequency bands and wider for high frequency bands.

For training, we chose a supervised approach, where we use as labels the output of anomaly detection from the accelerometer signal as ground truth. The features are normalized using their values at the start of the experiment, as the absence of normalization would be too sensitive to microphone gain and positioning.

There are several choices for the deep AI model, such as a deep neural network, a recurrent neural network, a temporal convolutional network or a transformer. In this work, our choice is a deep neural network (DNN), whose specifics are given in Section 4.3.

4. Experimental Setup

The bearing datasets used in this study are collected in Flanders Make's Smart Maintenance Living Lab (Ooijevaar et al., 2019). This lab is developed as an open test and development platform and aims to support the adoption of condition monitoring technologies in the industry. It consists of seven identical drive train sub-systems. The setups are designed to perform accelerated lifetime testing of bearings and run bearings to their end-of-life. The accelerated lifetime test allows to create surface fatigue faults in bearings and monitor the fault evolution and accumulation during the (accelerated) life.

4.1. Bearing test rig and accelerated lifetime experiments

One of these experimental setups to perform the accelerated lifetime test is shown in the middle image of Figure 8. The setup comprises of a single shaft with a test bearing. The shaft is supported by a support bearing on each side. The test bear-

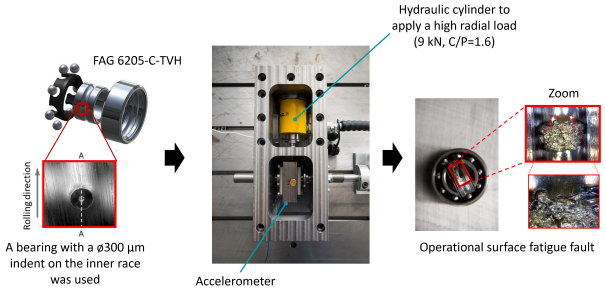

Figure 8. Illustration of the initial bearing state (left), the experimental test rig setup designed to perform accelerated life tests (middle), and the final state, a surface fatigue fault at the inner race of the bearing (right).

ing is lubricated by an internal oil bath. The setup is driven by a motor at a rotation speed up to 3000 RPM. In this work, we focus on experiments driven at 2000 RPM. Each setup is equipped with an accelerometer, temperature sensor, load sensor and speed sensor. The radial accelerations are measured at a sampling frequency of 50 kHz by an accelerometer attached to the bearing housing. The rotational speed and radial load of each setup can be controlled, such that each setup can operate at stationary and non-stationary operating conditions. An industrial Beckhoff control platform is used to acquire and store the sensor signals and to control the speed and load of each setup.

In total more than 70 bearing accelerated life tests have been performed on a FAG 6205-C-TVH deep groove ball bearing resulting in surface fatigue faults at the inner race. Two mechanisms are used to accelerate the bearing lifetime:

- A high radial load up to 9 kN (C/P = 1.6) is applied to the bearing outer ring with a hydraulic cylinder.
- Before the start of the test a small initial indentation of approximately 300 µm was created in the bearing inner race using a Rockwell C hardness tester. This indentation is used as a local stress riser and represents a local plastic deformation caused by, for instance, a contamination particle.

The accelerated life time tests are stopped as soon as 20g peak-to-peak accelerations are reached, resulting in severe rolling contact surface fatigue at the inner race (Halme & Andersson, 2009). The start and end condition of the inner race of one of the test bearing are shown in the left and right images of Figure 8.

4.2. Acoustic setup

The acoustic signals are acquired through two B&K 4189A21 microphones sampled at 50 kHz. One of them was placed under the safety cover of the bearing test setup, and the other outside of the cover, as showed in Figure 9. These microphones will be referred to respectively as *IntMic* and *ExtMic*

(a) Microphone inside the safety cover, referred to as *IntMic*.

(b) Microphone outside the safety cover, referred to as *ExtMic*.

Figure 9. Illustration of microphone positions used for the experimental recordings.

Figure 10. Illustration of the acoustic scene.

throughout the rest of the paper.

The experimental setup is situated in a large laboratory area at Flanders Make's facilities in Leuven, which has an uncontrolled and reverberant acoustic environment shown in Figure 10. It is a concrete room that contains many different kinds of setups such as drivetrains, looming machines etc., sometimes running simultaneously. There is also human activity with technicians and engineers running and maintaining the setups, or going about their daily activities. Due to the varying sizes of the setups here, sometimes small vehicles like forklifts or the crane integrated into the laboratory can also operate here. This makes the background noise potentially quite complex.

Specifically for the dataset we collected, there are a few common sources of noise that are often present, and we chose them as our focus for techniques that need us to characterize the kind of background noises that need suppressing. The most consistent, and arguably the simplest, disturbance is that coming from the hydraulic pump used to apply load on the test bearing. This pump activates roughly every minute in order to keep the pressure, and thus the load, constant. This cre-

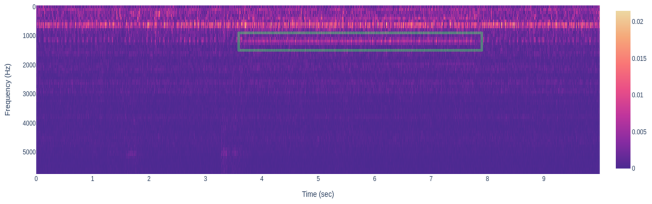

Figure 11. The spectrogram of an instance of pump activation with a faulty bearing being tested, captured by the *ExtMic* microphone. The pump activity is highlighted inside the green rectangle.

ates a very audible and consistent noise that is quite apparent in the signal, as can be readily observed in the spectrogram presented in Figure 11. The second common disturbance is a sharp, impulsive noise made by a nearby setup dropping metal pipes in a container. This happens once or twice a minute. Finally, on multiple occasions, there are people who are either walking or standing around the setup while in conversation. These are usually captured by the microphones, and constitute a third kind of disturbance we are interested in.

4.3. Processing configuration

The feature extraction is performed on segments of 10 seconds. For the BPFI feature, the tolerance of the SES peak search around the BPFI and its harmonics is set to 1.5% of the theoretical frequency (Eq. 1). The number of harmonics n_{harm} for the calculation of ξ_{BPFI} (Eq. 2) is set to 4. The mel spectrogram is calculated on 64 mel bands. The features are normalized using Z-score normalization, where the mean and standard deviation are computed on the first 30 minutes of the corresponding experiment. In experiments that lasted for more than one day, this calculation is done for each day.

Our dataset consists of two groups of run-till-failure experiments, all at 2000 RPM, where the bearing developed an inner race fault. The first group is characterized by continuous monitoring, where the sensors (accelerometer and microphones) constantly acquired data. This group contains 11 experiments, 7 of which run till failure and 4 of them ended prematurely. The second group contains 5 experiments were periodic monitoring of 1 second every 10 seconds was applied, 3 of them run till failure and 2 of them ended prematurely. This group is only used for training purposes. For cross-validation purposes, the dataset is split in three folds.

In addition to the captured data, a set of ground truth labels is also provided. It should be noted here that this labelling is not based directly on the physical state of the bearing, since it would not be available without stopping the test and dismantling the bearing, but based on analysis of the data captured by the accelerometer. Using this labelling the moment in time where the bearing starts having faulty behavior is determined. Data prior to this moment is then considered as healthy, and data afterwards is considered faulty.

For the deep AI model, we choose a deep neural network (DNN) with an input layer, three hidden layers of 32, 16 and 8 units, and an output layer.

5. Results and Discussion

In this section, we evaluate and compare the performance of both noise-aware and noise-unaware methods for bearing fault diagnosis using the microphone *ExtMic* signals, as its location outside of the safety cover of the setup is the most realistic. For clarity, we provide a summary of the methods evaluated in the following list.

1. **Noise-unaware methods**: These methods, described in Section 2.2, do not take the presence of noise explicitly into account. A diagram illustrating both methods is shown in Figure 4.
   (a) **Baseline**: The baseline method is based on the BPFI feature with median-smoothing.
   (b) **Shallow AI**: This method uses the BPFI and RMS features as input to a two-layer fully connected NN (shallow AI) and median-smoothing to achieve a diagnosis result.
2. **Noise-aware methods**: These methods aim to increase their robustness to noise, as explained in Section 3.
   (a) **Noise-aware smoothing**: The method described in Section 3.1, where the weights of the smoothing filter depend on the detected noise level. Its diagram is shown in Figure 5.
   (b) **Deep AI with hybrid features**: The method detailed in Section 3.2. It combines the BPFI and RMS features with mel spectrum features, where the latter are the input of the DNN (the deep AI model). It uses median-smoothing to obtain a diagnosis result. Its diagram is shown in Figure 7.
   (c) **Deep AI combined with noise smoothing**: This method is a combination of the two previous methods, i.e., the methods 2a and 2b.

5.1. Performance metrics

We use several metrics to assess the diagnostic performance of the proposed methods.

- **EPR**: The point on a precision vs. recall plot where these two metrics are equal. A high score indicates a high ratio of true positives with respect to both predicted positive samples and real positive samples. Expressed as a percentage.
- **ROD**: Rate of detection, the rate at which faults are detected before the safety stop, at a given precision value, equivalent to *recall*. In our case, we compute this value at 99% precision. More formally, if TP and FN denote respectively the true positive and false negative counts,

then

$$ROD = \frac{TP}{TP + FN}, \tag{3}$$

where $\frac{TP}{TP+FP} > .99$. Expressed as a percentage.

- **TOFD**: Average time of first detection in seconds, i.e, the time between the occurrence of the fault and the first time the model detects its, for a given precision value. In our case we compute this value at 99% precision, and it is only considered if the ROD is 100%. In the same way as our ground truth labels, the occurrence of the fault is defined based on accelerometer data. Note that this metric can be defined w.r.t. any source more reliable than the acoustic signal we are using. Formally, this metric is computed as

$$\frac{1}{N}\sum_{n=1}^{N} t_{fault,n} - t_{detection,n}, \tag{4}$$

where $ROD = 1.0$, $t_{detection,n}$ is the time of detection, and $t_{fault,n}$ is the time the fault occurred, both for the n-th experiment.

Regarding the TOFD metric, note that there is always some delay between the occurrence of a fault and its detection. There are two main sources of this delay. The first one is related to smoothing, and is not affected by the fact that we are running accelerated lifetime tests. This means that it would remain constant (for a given smoothing filter) even in regular testing. The second is the time gap between the fault being available to a vibration sensor versus it being available to an acoustic sensor. This delay pertains to the evolution of the fault, and therefore scales with the fault accelerations applied during the testing procedure. If we were to run regular lifetime tests, these delays would be multiplied by a corresponding factor. This means that, while the first delay is adjustable, the second delay is a consequence of the physics of the system and can only be reduced so much. It is a hard constraint on acoustic monitoring.

5.2. Performance results

The performance metrics for our evaluated methods are displayed in Table 1.

5.2.1. Performance of noise-unaware methods

It can be clearly seen that the baseline method's performance is quite poor, since its EPR is barely over 50 %, and at 99% precision it is only able to detect 14% of the faults.

The introduction of the shallow AI causes a significant jump in model performance, where it can now reach 100% ROD at a precision level of 99%, and an EPR point of 69%. However, note that the TOFD of 1520 seconds, about 25 minutes, is quite high, due to the required size of the smoothing window to reach a high precision. An example from the health indicator obtained by the shallow AI method is displayed in Figure 12, both before and after median smoothing. It can be observed that the shallow AI allows for a clear distinction of the faulty state when the fault actually develops, but it is the smoothing that removes the high amount of false positives. However, each process introduces a noticeable delay in the health indicator, which is expected to be dependent on the particular characteristics of the background noise.

Table 1. Performance results of the evaluated methods.

Method	EPR	ROD	TOFD
Baseline	55%	14%	-
Shallow AI	69%	100 %	1520 sec
Noise-aware smoothing	72%	100 %	1031 sec
Deep AI & hybrid features	79%	100 %	605 sec
Deep AI & hybrid features + noise-aware smoothing	81%	100 %	600 sec

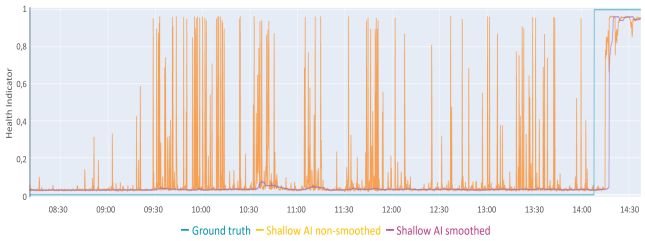

Figure 12. Example of health indicator obtained from the shallow AI method before and after median smoothing.

5.2.2. Performance of noise-aware methods

The use of noise-aware smoothing increases the EPR to 72%, maintains the ROD of 100%, and its TOFD is decreased by 33% with respect to the TOFD of the shallow AI method. This is a strict but moderate improvement over the best noise-unaware method.

The further addition of the mel spectrum features and the deep AI model causes a significant leap in performance, where the EPR point reaches 79%, and the TOFD is decreased by 42% and 60% of the TOFD values of the noise smoothing and shallow AI methods respectively. This improvement demonstrates the ability of the DNN to learn complex patterns from the mel spectrum features, and to complement the BPFI feature to achieve a better diagnostic performance. An example from the health indicator obtained by this method is displayed in Figure 13, both before and after median smoothing. It can be readily observed that before smoothing, the false positives are notably less frequent than in the example of the shallow AI method from the same experiment, shown in Figure 12. Smoothing removes these false positives, but crucially it in-

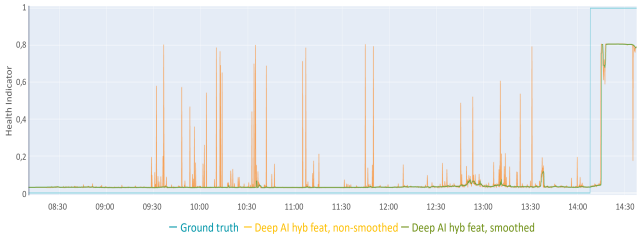

Figure 13. Example of health indicator obtained from the deep AI with hybrid features method before and after median smoothing.

Table 2. EPR points for median and noise-aware smoothing for methods 1b, 2a, 2b and 2c. Hybrid refers to the combination of BPFI and RMS with mel spectrum features.

Smoothing type	Features	
	BPFI & RMS	**Hybrid**
Median	69%	79%
Noise-aware	72%	81%

troduces less delay than in the shallow AI example. The reason is that, as the deep AI method produces fewer false positives, the smoothing window can be shorter, thus minimizing the additional delay introduced by this step.

Last of all, the combination of the two noise-aware methods, i.e., where noise-smoothing is applied to the deep AI method with hybrid features, achieves a moderate increase of the EPR until 81%, and no significant reduction in TOFD.

5.3. Effect of smoothing: noise-aware vs median

Note that the inclusion of noise-aware smoothing results in a moderate EPR improvement over the same method using median smoothing. This can be seen in Table 2, where the EPR points are arranged depending on the smoothing type and features that methods utilize. The reason is that noise-aware smoothing addresses the following issues:

- For a true increase in the anomaly score, noise-aware smoothing is typically faster to respond due to its non-even weighting.
- In case of quickly fluctuating anomaly scores, median smoothing is a lot less stable due to its inability to choose what to prioritize.

In most other cases noise-aware smoothing behaves comparably to median-smoothing, offering the same benefits. This makes noise-aware smoothing an attractive enhancement, although it comes with the additional cost of designing appropriate noise detectors for the acoustic scene where the machinery of the monitored bearings operates.

6. Conclusion

In this study, we have focused on developing robust methods for acoustic condition monitoring of inner race faults in rolling element bearings in industrial environments. This is a challenging problem due to the strong influence of background noise, which introduces a high amount of false positives and delays fault detection, resulting in poor diagnostic performance. Our two proposed noise-aware methods have different levels of complexity. The first and simplest one is noise-aware smoothing, which adapts the smoothing weights according to the detected noise levels. The second and more complex one is a deep AI model that uses mel spectrum features and acts in combination with the bearing fault frequencies to achieve a diagnostic result. These methods have been trained and tested with an accelerated bearing lifetime dataset acquired in the Flanders Make Smart Maintenance Lab, which is a reverberant environment where strong and diverse acoustic disturbances were present.

The results demonstrate significant improvements over the noise-unaware baseline, both in diagnostic performance and in detection time, using a single microphone signal. Moreover, these benefits are distinct both when the noise-aware methods are applied independently or in combination, so they can thus be chosen according to the monitoring requirements of each particular use case. In summary, we have shown that, when employing adequate strategies to increase robustness to noise, acoustic monitoring can be a cost-effective and practical alternative for vibration monitoring. Future work in this problem involves studying the influence of the training dataset size on accuracy, applying and testing our strategies to other bearing fault types, and studying the effect of data augmentation in the training of the deep AI model.

Acknowledgment

This research work was supported by Flanders Make, the strategic research centre for the manufacturing industry, and more precisely by the ACMON ICON research project. Authors would like to thank this project for funding the research presented in this paper.

References

Cardenas Cabada, E., Leclere, Q., Antoni, J., & Hamzaoui, N. (2017). Fault detection in rotating machines with beamforming: Spatial visualization of diagnosis features. *Mechanical Systems and Signal Processing*, *97*, 33-43. (Special Issue on Surveillance)

Glowacz, A. (2019). Fault diagnosis of single-phase induction motor based on acoustic signals. *Mechanical Systems and Signal Processing*, *117*, 65-80.

Halme, J., & Andersson, P. (2009). Rolling contact fatigue and wear fundamentals for rolling bearing diagnostics -

state of the art. *Journal of Engineering Tribology*, *224*, 377–393.

Lee, J., Wu, F., Zhao, W., Ghaffari, M., Liao, L., & Siegel, D. (2014). Prognostics and health management design for rotary machinery systems—reviews, methodology and applications. *Mechanical Systems and Signal Processing*, *42*(1), 314-334.

Mian, T., Choudhary, A., & Fatima, S. (2022). An efficient diagnosis approach for bearing faults using sound quality metrics. *Applied Acoustics*, *195*, 108839.

Ompusunggu, A. P., Devos, S., & Petre, F. (2013). Stochastic-resonance based fault diagnosis for rolling element bearings subjected to low rotational speed. *International Journal of Prognostics and Health Management (IJPHM)*, *4*.

Ooijevaar, T., Pichler, K., Di, Y., Devos, S., Volckaert, B., Hoecke, S. V., & Hesch, C. (2019). Smart machine maintenance enabled by a condition monitoring living lab. *IFAC-PapersOnLine*, *52*(15), 376-381. (8th IFAC Symposium on Mechatronic Systems MECHATRONICS 2019)

Rabiner, L., & Schafer, R. (2010). *Theory and applications of digital speech processing* (1st ed.). USA: Prentice Hall Press.

Randall, R. B. (2011). *Vibration-based condition monitoring: Industrial, aerospace and automotive applications*. John Wiley & Sons, Ltd.

Randall, R. B., & Antoni, J. (2011). Rolling element bearing diagnostics—a tutorial. *Mechanical Systems and Signal Processing*, *25*(2), 485-520.

Ricardo Mauricio, A. M., Denayer, H., & Gryllias, K. (2022). Time-domain beamformed envelope spectrum of acoustic signals for bearing diagnostics. In *Conference proceedings of ISMA 2022 - USD 2022.*

Ricardo Mauricio, A. M., Denayer, H., & Gryllias, K. (2023). Beamformed envelope spectrum of acoustic signals for bearing diagnostics under varying speed conditions. In *Proceedings of NOVEM 2023.*

SileroTeam. (2021). *Silero models: pre-trained enterprise-grade STT / TTS models and benchmarks.* `https://github.com/snakers4/silero-models`. GitHub.