

Reviewer 1 and authors' response

Summary

The paper compares envelope-spectrum condition monitoring against an autoencoder reconstruction z-score (AI method) as an index for monitoring the health of bearings under fatigue tests. The conclusion notes comparable detection performance in three of the five cases and better performance of the AI method in the others.

The authors would like to thank the reviewer for their work and answer the comments individually in the following.

Scientific significance:

- The research question is well chosen. Building a single index for monitoring bearing health is very valuable.

Thank you!

- The AI method is not novel. It was already developed in the authors' own prior work. I would therefore classify this paper as a benchmark study rather than a methodological contribution.

Agreed, this paper describes how the methods are applied to a set of data.

- As a benchmark study, I think the test size is small. A single dataset with only five test cases, three of which show neck-and-neck performance, is not representative enough.

The authors agree that the test size is rather small. The original data set includes several more tests. These, however, have significantly different conditions (mechanically pre-conditioned bearings) which explanation would add several aspects that are not related to the core story. On the other hand, several more recent works in the field of bearings testing rely on a equally small number of samples. This is simply a consequence of time and budget restrictions. Publications with similar numbers of bearings per test conditions include:

Schneider et al. (2024) conducted fatigue testing of bearings with electrical erosion damage of differing levels. The bearing testing done at two of those levels consisted of 4 and 6 failures, with 0 and 2 tests stopped respectively before failure had occurred. (Schneider, V., Krewer, M., Poll, G., and Marian, M.: Effect of Harmful Bearing Currents on the Service Life of

Rolling Bearings: From Experimental Investigations to a Predictive Model, Lubricants 2024, Vol. 12, Page 230, <https://doi.org/10.3390/LUBRICANTS12070230> , 2024).

Dinkel et. al (2020) conducted fatigue testing of bearings with artificial indentations made by hardness testers with differing levels of load applied on the hardness testing equipment. One of the sets of bearing testing data at one of the levels of load applied consisted of 4 failures with 7 tests which were stopped before failure had occurred. (Dinkel, M., Zheng, X., Warmuth, M., and Correns, M.: Surface Damage in Rolling Bearings and the Impact on Rolling Contact Fatigue, in: ASTM Special Technical Publication 1623, pp. 147–168, ASTM International, ISBN 9780803176928, <https://doi.org/https://doi.org/10.1520/STP162320190094>, 2020)

Kehl (2020) conducted fatigue testing of bearings with roller-impact indentations. One of the sets of tests contained 6 failures and 18 tests which were stopped before failure had occurred. (Kehl, J. H.: Ermüdungslebensdauer bei Oberflächenbeschädigungen, <https://fva-net.de/aktuelles/detail/ermuedungslebensdauer-bei-oberflaechenbeschaedigungen-522>, 2020)

Scientific quality:

- Lines 237–238 state that "the mean and standard deviation calculated from the reconstruction error of the healthy training dataset D after the completion of the training." I infer that μ and σ are computed from the training data itself. This is not ideal; a better approach would be to hold out a slice of healthy data for validation rather than using it in training.

The authors would not expect a significant difference in the results when using the *validation data, as both the training and validation datasets consist of healthy data and follow the same distribution. The mean and standard deviation calculated for one experimental dataset are provided below:*

Training: mean: 1.8560874998821142e-05, std:2.0925273043916233e-05

Validation: mean:1.863536749466268e-05, std: 2.1005096040126406e-05

The differences in both the mean and standard deviation between the training and validation datasets are less than 0.5%, indicating that the two datasets have very

similar statistical characteristics. Therefore, using the validation data instead of the training data is not expected to have a meaningful impact on the results.

Furthermore, during network training, a portion of the validation data is used to monitor the model's convergence and assess its performance on both the training and validation datasets.

- The AI method's loss function is reconstruction-based, meaning it treats all indicators equally and aims to minimize reconstruction error across all of them. However, some indicators may be noisy and dominate the high-level indicator, while others may contribute to the reconstruction error without being informative of the fault. I would suggest an ablation study to identify which indicators help the z-score raise a timely alarm and which ones degrade its performance.

Different condition indicators are sensitive to different fault mechanisms, and no single indicator can reliably detect all fault types. Consequently, the proposed framework combines a diverse set of complementary indicators to capture fault-related information from multiple sources. Therefore, an indicator that does not contribute to a specific fault-detection case should not be considered uninformative, as it may be highly sensitive to other fault mechanisms.

- The thresholds are not consistent: the classical approach uses 10·Iref, while the AI method uses 5/10. The paper states these were empirically tuned. What is the false-alarm rate associated with these thresholds?

The actual fault alarms occur well above the normal baseline; therefore, we intentionally kept the threshold levels relatively high to minimise false alarms, since stopping the experiment, opening the bearing to find no damage and then restarting it can influence its result. The threshold values were not formally tuned; instead, they were selected based on observations from the experimental results and prior expert knowledge, reflecting the practical preference to avoid unnecessary inspections, particularly for offshore wind turbines. While a higher threshold may delay fault detection, it helps prevent unnecessary shutdowns caused by false alarms.

Presentation quality:

- Just below the caption of Figure 11: "Disclaimer. The authors disclaim this work to be a very good paper." This appears to be a mistake.

We have removed this sad excuse of a joke.

- Typo: Section 4 heading "Discussion and **Conlusions**."

Thanks, corrected.

- Table 1, row "One-class anomaly": the citation appears as "(?)".

In our source file, the citation is correct and displays as well in the pdf file we compile from it. Unsure how to fix it at this stage, we leave it the way it was but are sure to keep an eye out for it for the final review by WESJ.

Recommendation:

- Moderate revision. The AI method requires further clarification. Some sentences in Section 1.3 are long and hard to follow.

The clarification of the AI methods is part of the individual comments. Long sentences in 1.3 have been revised and split up.

The false-alarm rate of the thresholds, and a consistent approach for choosing them, need to be discussed.

See above answers to the respective comments.

In the abstract, the claim of an "even more pronounced advantage for enhanced monitoring in more complex machines like wind turbines" is not numerically supported by the experiments. I would suggest framing it as a hypothesis for future work.

Thank you for this suggestion, we have changed the abstract to:

warnings were significantly earlier and clearer. As classical monitoring benefits most from the simplified test environment, it is reasonable to assume an even more pronounced advantage for enhanced monitoring in more complex machines like wind turbines. *The verification of this assumption is part of future work.*

Reviewer 2 and authors' response

The article compares the performance of two condition indicators derived from vibration measurements acquired during accelerated tests of Ø100 mm thrust bearings. The first condition indicator is a more traditional measure based on 2x the ball spin frequency from the Hilbert transform of the vibration, while the second is the z-score of the reconstruction error from a multivariate autoencoder of a number of similar condition indicators. The use of deviation of the reconstruction error from typical levels for a healthy bearing as a composite health indicator is interesting, and in general it performs better than the more traditional method in 4 of the 5 tests. I do like the article and recommend it for publication; however, I have a few general and several specific comments and questions for the authors to consider.

The authors would like to thank the reviewer for their work on the paper. The review was clear and easy to follow, and the authors highly appreciate how the reviewer even suggested solutions in many places which made implementation even easier. In the following, we are to reply to each comment individually.

General

- It took quite some time for me to understand what is often described as 2 or more AI methods. This is true of the title (“AI enhanced fault indicators”, emphasis on plural), abstract (“two ways of using AI”), general descriptions in the article, and Conclusions (“two approaches”). I found this confusing and rather prefer the research question in Section 1.4 (“Can a combination of normal behavior model and high-level health indicator”, emphasis on combination). I recommend the authors make these descriptions more consistent. Really to me, I see what I described in my introduction to my review: the research value lies in using the z-score of the reconstruction error from a multivariate autoencoder of a number of vibration-based condition indicators. That is 1 method. At the end of the day, a single condition indicator (the “high-level indicator”) from the autoencoder is compared to a single traditional condition indicator (the “RPBF amplitude from HT+FFT”).

This is a very interesting remark. To the authors, it is quite obvious that the autoencoder and the congregation to one high/level indicator are two different things which apply different methods, and they also consider this to be clearly separated in the work.

An example from the introduction:

(et al., 2023). In the context of this work, such AI techniques are in two fields:

1. Combination of individual indicators to high-level indicators
2. Use of normal behavior models to determine faults without knowing their signal characteristics

Which is then followed by one paragraph of literature for high-level and more for autoencoder or other methods on single signal (bands)

However, it occurred to the authors that this lacks some clarity on what exactly ‘individual indicators’ are, so we rephrased:

1. Combination of individual ~~frequency bands~~~~indicators~~ to high-level indicators
2. Use of normal behavior models to determine faults without knowing their signal characteristics

While this work uses high-level indicators to combine information of various frequency bands, the underlying methods can also be applied to aggregate different types of indicators like different sensor types. Despite the wide use of vibration analysis for ~~CM~~~~condition monitoring~~, and its potential in practical application, the fusion of multiple condition indicators into a single interpretable metric remains relatively underexplored in the literature. Existing work includes a regression-based normal behavior modeling approach that labels individual condition indicators as healthy or faulty, which are then combined to

We have also tried to include a hopefully helpful remark in the description of Figure 6 (first figure of the results):

bearing 2. The left part of the figure shows the curves of the classical condition monitoring of RPBF in the top graph and the high-level indicator ~~which combines all frequency bands~~~~of the AI-enhanced monitoring~~ in the lower graph. The classical

- With that understanding (if indeed it is agreed upon), I don’t believe the manuscript described exactly which traditional condition indicators were used in the autoencoder. That is, which formed $\mathbf{X}_i$ in Equation (9). I understand this may be considered proprietary, but if I could wave a magic wand this would be described better.

Figure 5 gives an overview of the entire process, starting from the raw vibration signal over the first stage which splits it into frequency bands to the second stage which applies statistical indicators. The text then lists the following as indicators: six statistical indicators, root mean square, crest factor, kurtosis, number of 4-sigma outliers, time negentropy, and spectral negentropy (Line 208 and 209). Expelliarmus.

- In the Title and abstract much is made of AI. I will admit that I am not an expert in multivariate encoders (I did Google it), but my understanding of them is that rely on machine

learning and/or neural networks. So, the “we must say AI” feels a bit much. Again with the understanding (if it is correct) that the research value lies in using the z-score of the reconstruction error from a multivariate autoencoder of a number of vibration-based condition indicators, I would just say that.

The authors fully agree. Thankfully, recent political developments have replaced it with a much better, a really great, new acronym. Anyhow, we found it tremendously to replace AI but for two instances. This is because ‘AI-enhanced’ in this work would translate to ‘combination of multivariate autoencoder and high-level indicator’ which is super-precise, but also very bulky. Even in the very title and after much pondering, it seems to like a very bulky term to replace AI-enhanced.

- The idea of verification is used twice in the abstract and once in the conclusions. I agree the bearing tests were quite nice, but fully “verifies” I think is a bit much. The “example results” in the Title is more honest.

Yes, precisely. We have changed it accordingly.

- Combining several of these comments together the Title I think is more honestly something like “Autoencoder-based vs. classical vibration-based bearing condition monitoring: example results of bearing tests”.

Yes, this sounds like a better title than the current version, but it omits the high-level indicator aspect, so we opt for “Autoencoder and high-level indicators vs. classical vibration-based bearing condition monitoring: example results of bearing tests”.

Further, in the abstract lines 5 to 7 is more like “...less manual interaction. Because of the need for less supervision, this paper develops a single condition indicator from the reconstruction error of a number of vibration-based condition indicators using a machine learning-based multivariate autoencoder and compares its performance to a traditional single vibration-based condition indicator over 5 accelerated tests of Ø100 mm thrust bearings. In these bearing tests, the autoencoder-based monitoring...”

In the light of previous answers, this is how we rephrased it:

for enhanced methods that require reduced supervision and less manual interaction. The present work ~~applies~~^{verifies} two ways of using artificial intelligence that ~~have the potential to~~^{fulfill} this requirement. These are normal-behaviour models and high-level indicators. ~~They are applied to~~^{The verification includes} test data ~~analysis~~^{of accelerated} small-scale bearing tests of Ø100 mm thrust bearings ~~and considerations of transfer to wind turbines. The combination of normal-behaviour models and high-level indicators~~^{In bearing tests, enhanced monitoring} gives comparable or significantly earlier warnings than classical ~~condition~~^{condition} monitoring. In three of five performed tests, the warning thresholds were passed at comparable times. In two, the

- Other than the mention of the need for testing on more complicated components like actual gearboxes, I'm not sure the "transferability to wind turbines" is really discussed in the manuscript. I'm happy to have the article submitted to Wind Energy Science; however, the meat of it is suitable for many publications such as MSSP, etc. Given that, I'd recommend "transferability to wind turbines" be removed from the title and "considerations of transfer to wind turbines" be removed from the abstract as they simply aren't discussed. Wind turbines are just one of many applications of this method.

The authors agree. We have mentioned a few aspects of the transferability and future works will contain more on this, but it is not touched substantially in the present work.

Introduction

- Line 25: My understanding of clause 7.6.1.3 in IEC 61400-1 is that it only allows a lower partial safety factor of $\gamma_n = 0.9$ for replaceable bearings in a structural ultimate limit state analysis. So, it just feels like this statement "IEC 61400-1 foresees reduced safety factors for monitored bearings" is a bit misleading. For example, the bearings discussed in the current article are most similar to bearings in the drivetrain (main bearing, gearbox, and/or generator), rather than part of the structure. Can the authors correct me if I'm wrong, or further explain what they mean? Another point to consider is that although IEC 61400-4 requires some bearings to be monitored, it does not include any special "credit" for doing so to my knowledge.

First, the authors would like to acknowledge their chosen wording is short and misleading, as the reviewer suggests.

We have tried to find our way around IEC61400-1 Chapters 7 (structures) and 9 (Mechanical Systems). The definition of a 'Mechanical System' is 'any system which does not consist solely of static structural components [...] but transmits relative motion [...]', which, in the eye of the authors is a bearing. Chapter 9 then specifies that partial safety factors are chosen according to component class 2 of 7.6.1.3, and that equals a factor of 1. Now, the authors could be funny and argue that this is reduced in comparison to class 3 with factors of 1.2, but we would rather not go down that road. In the light of this, the statement should be omitted.

*But then the examples for class 1 in 7.6.1.3. are 'replaceable bearings', as the reviewer remarked correctly. Now it escapes the authors completely how a bearing can be mentioned outside of Chapter 9 because this implies the existence of a bearing which does *not* transmit relative motion and the authors have completely missed the lectures on*

those in their studies. Putting aside the aspect of relative motion: especially the bearing which are part of the structure like pitch, yaw, and main, are hard to replace.

But say someone wants their bearings which calculation returns a lower expected fatigue life than the turbine life to pass the certification, they would most likely go along and argue exactly these are 'replaceable bearings' (aren't all bearings?) and blissfully ignore the words of Chapter 9.

Thoroughly confused, the authors deleted the statement on reduced safety factors and got another coffee.

- Line 65: What does “monitor multiple indicators of different scales” mean in this context? In particular, what is meant by “scale”? I believe it refers to condition indicators with different units, for example an RMS versus a particular vibration amplitude (e.g. BPFO)? If this is the right sense, I might recommend simply “monitor indicators from different techniques”. The word techniques referring back to the 4 techniques described earlier in the section.

Marvellous suggestion, we are happy to revise it accordingly:

uous monitoring to assess the condition of mechanical components. Early fault detection becomes increasingly challenging when experts manually monitor multiple indicators from different postprocessing techniques, and subtle changes in a trend can easily be overlooked. Furthermore, it is essential to track multiple indicators because different indicators are sensitive

Results

- It is true that the indication is given earlier by the autoencoder-based method. However, in such an accelerated test it is difficult to understand how much earlier it is, especially since the x-axis does not begin at 0 in Figures 6 to 10. Maybe a more interesting metric to graph on the x-axis instead of test time would have been the proportion of L10 consumed? Further, a table summarizing the difference in time, possibly in %, across all 5 tests would also be helpful. I was flipping back and forth between the different figures several times in order to understand the difference in time, typically in fractions of a day except of course for the 2 tests where the traditional method gave no indication at all. Honestly, I'd focus on those 2 tests as being more significant than the 2 others where the auto-encoder was marginally better (and still be honest about the 1 test in which it wasn't).

We agree that a quantitative comparison would help make the results clearer. We have added a column to Table 5 showing the test duration as a percentage of the calculated L_{h10} life, which should make it easier to compare the tests. Regarding Figures 6–10, we would prefer to retain the current x-axis ranges, as showing the entire test duration would compress the signals and make it more difficult to observe their development.

Discussion and Conclusions

- Line 313: Here a change in bearing friction is mentioned as a cause of the change in bearing frequencies. I'm not sure this can be concluded, or even if it could why it matters? I'd recommend deleting this portion of the sentence.

The vein of thought behind this very brief statement

“In one case, a notable change in the frequencies was detected, most likely due to a change in bearing friction. “

is: If the bearing friction increases, it is likely for gross slip of the rolling body set to appear. The rolling speeds and with them the frequencies such as BPFO will be reduced by this. If a static algorithm monitors a fixed frequency band, it is possible that an initially detected damage at some points leaves the boundaries of this band because the rolling body set moves significantly slower than its kinematic speed. The damage indicator in this scenario would first increase and then decrease.

The authors would thus argue such a behavior to be of certain relevance in the postprocessing of the data and have rephrased as follows:

In ~~T4one~~-case, a notable reduction in the frequencies was detected over the test time. This frequency change implies a gross rolling body set slip, which can be caused by increased friction, ~~most likely due to a change in bearing friction.~~ The
315 classical CM ~~postprocessing~~present-work applied an automated approach to track such frequency changes by identifying the

Minor comments:

- Line 4 and 34: It's arguable the extent to which commercial systems “include” or “require” significant human supervision and interpretation of measurement data. I'd feel a bit better if this were softened a bit to “often include” or “often require”, but I'm not gonna battle about it.

Neither are we – ‘often’ is fine for us:

for maintenance planning and in-situ counter-measures and thus reduces operational costs. Current commercial methods ~~often require~~~~include~~ significant human supervision and interpretation of measurement data. Larger fleets of assets raise the need for

- Line 23: I think this sentence reads better if the latter half would be “...major repairs in wind turbines, continuous health monitoring of bearings facilitates more efficient maintenance thereby reducing operations and maintenance costs”.

Absolutely, thank you.

Given the costs for downtime and the necessary logistical planning for major repairs in wind turbines, continuous health monitoring of bearings facilitates more efficient maintenance thereby reducing operations and maintenance costs.,~~it is but~~
25 ~~evident that continuous health monitoring of bearings in wind turbines reduces operations and maintenance costs, because it~~
~~facilitates more efficient works~~ IEC 61400-1 foresees reduced safety factors for monitored bearings (IEC, 2019). The use of

- Line 27: I believe “ISO28016-21” should be “ISO 20816-21”.

Thanks for spotting this.

vibrational sensors and analysis of their data is one c
is outlined in ISO~~20816~~~~28016~~-21 (ISO, 2025), but f
et al. (2025) give a comprehensive overview of the st

- Line 41: CM is defined as the acronym for condition monitoring here; however, it is only infrequently (6 times) used in the remainder of the article whereas condition monitoring is used much more often (33 times). It is best practice to use the acronym consistently, or not at all.

You are very right. We have changed several instances of condition monitoring to CM throughout the text:

Hart et al. (2023) analysed large-scale field data on main-bearing failures, demonstrating the widespread application of classical ~~CMecondition-monitoring~~ in wind turbine drive trains and its importance for maintenance planning. For a comprehensive methodological background, Randall and Antoni (2011) provide a tutorial summarizing classical signal-processing techniques such as FFT, STFT, and envelope analysis, and discuss best practices for ~~CMecondition-monitoring~~ of rolling-element bearings.

~~CMeondition-monitoring~~ indicators track distinct signal patterns to diagnose different fault mechanisms. They require continuous monitoring to assess the condition of mechanical components. Early fault detection becomes increasingly challenging when experts manually monitor multiple indicators of different scales, and subtle changes in a trend can easily be overlooked. Furthermore, it is essential to track multiple indicators because different indicators are sensitive to different fault types. These limitations challenge ~~classical CMtraditional-condition-monitoring~~ for large fleets of complex machines, such as wind turbines. Artificial intelligence (AI) techniques can simplify and automate the ~~CMeondition-monitoring~~ process (Chesterman et al., 2023). In the context of this work, such AI techniques are in two fields:

Despite the wide use of vibration analysis for ~~CMeondition-monitoring~~, and its potential in practical application, the fusion of multiple condition indicators into a single interpretable metric remains relatively underexplored in the literature. Existing

A regression-based method trained solely on vibration-based ~~CMeondition-monitoring~~ indicators, without using operating condition data, has been used to detect gearbox bearing faults (Rao et al., 2025). Also SCADA data collected during healthy operation has been used to train autoencoders for drivetrain health monitoring (Chen et al., 2021; Zhao et al., 2018). Table 1 summarises normal behaviour approaches from the literature.

Table 1. Summary of existing normal behaviour ~~CMeondition-monitoring~~ approaches.

- Line 42: The citation should be changed to “(ISO, 2015; ISO, 2016)”.

Indeed.

Classical condition monitoring (CM) in the context of this work is a physics-based approach that relies on vibration measurements and signal processing with static algorithms, see also ISO 13373-2 and ISO 13373-3 (ISO, 2016, 2015). The main objective is to identify changes in machine behaviour that indicate the onset and development of faults in one or more components.

- Line 55: A space is needed for “outer (BPFO)”, while “and” is needed at “respectively, and ball spin frequency (BSF)”.

Yes.

These methods rely on predefined characteristic frequencies of the drivetrain, such as the shaft rotational frequency, gear-mesh frequencies, and bearing defect frequencies such as ball passing frequency outer (BPFO) and inner race (BPFI), respectively, and ball spin frequency (BSF) (SKF, 2011)). Note the terms for characteristic frequencies commonly refer to ‘ball’ but are

- Table 1: A (?) is listed for a citation. I believe it should be Laouti et al., 2011. Additionally, DVSD is not defined elsewhere. Lastly, what is a “torque actuator”?

Remarkable that the reviewer even gives the solution for that lacking citation. However, in our source file, the citation is correct and displays as well in the pdf file we compile from it. Unsure how to fix it at this stage, we leave it the way it was but are sure to keep an eye out for it for the final review by WESJ.

DVSD refers to (VUB) ...

- Line 134 and 143: I recommend making these sentences part of the previous paragraph, rather than being their own paragraphs.

Absolutely, we changed it accordingly.

- Line 160: The BPFO, BPFI, BSF and I believe FTF were previously defined, so they do not need to be defined again here.

We did not find the FTF definition but agree on the earlier mentioning of BPFO and BPFI. Revised text at line 160 looks like this:

The resulting envelope spectrum provides clear signatures of the characteristic bearing defect frequencies for a rotating inner ring: ~~ball-pass-frequency-outer-race (BPFO,), ball-pass-frequency-inner-race (, BPFI)~~, fundamental train frequency (FTF), cage frequency), ball spin frequency (BSF), and their harmonics (McFadden and Smith, 1984). These theoretical frequencies calculate as per:

- Line 176: I believe alpha should be 90 here, not 0.

Many apologies, and please do not tell our fellow bearing engineers about this mishap as we would never ever be invited to any bearing conferences again.

- Lines 183 to 191: Would be better as a single paragraph, rather than 3 paragraphs.

Yes, that makes sense and has been changed accordingly.

- Line 186: Would read better as “Example influencing factors include the background noise...”

So it does.

is based on experience in operation and can differ for other monitored bearings and monitoring systems. ~~Example influencing factors include the background noise~~
~~Influencing factors are, i.e. the background noise~~ of a complex machine or the sampling frequency of the data acquisition. Step 2 involves an additional evaluation of the peak frequency in the frequency band to

- Line 227: Would read better as “The normal behaviour multivariate autoencoder trained on healthy...”

Significantly better.

225 The normal behaviour multivariate autoencoder trained ~~The trained normal behaviour multivariate autoencoder~~ on healthy condition indicators exhibits a bias towards reconstructing them with minimal error. As a result, the model struggles to recon-

- Line 285: Hereafter RPBF is italicized, whereas before it is not. I don’t believe it needs to be.

Neither do the authors, and they discussed this while writing, concluded to change all italic version to normal font and then overlooked these three instances. Thanks for spotting them.

Figure 11 depicts a waterfall plot of test T1. This plot shows frequencies from 1 to 50 Hz. The shaft frequency f_r is at 4.167 Hz and the ~~RPBF~~ is at 29.35 Hz. A clear amplitude increase at ~~RPBF~~ is observed closer to the end of the test and side-bands to ~~RPBF~~ gradually appear at $t = 32.4$ days. This matches the changes observed in Figure 6, which
285 contains only this frequency in the classical CM graph.

While Figure 11 basically confirms the information contained in Figure 6, the waterfall plots of tests T4 and T5. Figures

Change recommended:

et al., 2023). In the context of this work, such **AI** techniques are in two fields:

1. Combination of individual **frequency bands** **indicators** to high-level indicators
2. Use of normal behavior models to determine faults without knowing their signal characteristics

I would change this to:

In the context of this work such techniques are:

- Combining numerous condition indicators into a single high-level health indicator.
- Using normal behavior models to determine faults without employing fault signature.

Recommended Title:

AI-based high-level health indicator vs. classical bearing monitoring - example results of bearing tests.

(We are using an AI technique to combine condition indicators into a single high-level health indicator so it is basically comparison of AI-based automatic vs manual classical condition monitoring,)