Early Fault Detection in Rotating Machinery via Multivariate Autoencoder-Based Indicator Fusion

Faras Jamil[1], Nikhil Sudhakaran[2], Xinrun Liu[3], Matthias Stammler[4], Asger Bech Abrahamsen[5], Cédric Peeters[6], and Jan Helsen[7]

[1,6,7] *Acoustics & Vibration Research Group/OWI-Lab, Vrije Universiteit Brussel, Brussels, Belgium*

[1] *Artificial Intelligence Lab Brussels, Vrije Universiteit Brussel, Brussels, Belgium.*

[6,7] *Flanders Make@VUB, Flanders Make, Lommel, Belgium.*

[2,3,4,5] *DTU Wind and Energy Systems, Technical University of Denmark, 4000 Roskilde, Denmark.*

[4] *Fraunhofer Institute for Wind Energy Systems, Large Bearing Laboratory, Am Schleusengraben 22, 21029 Hamburg, Germany.*

ABSTRACT

This research proposes a normal behaviour multivariate autoencoder, which fuses multiple condition indicators into a single high-level health indicator to provide a comprehensive overview of the mechanical component's health. The model is trained exclusively on healthy data to learn normal behaviour and detect faults by observing deviations from this learned normal behaviour. The proposed method is validated in real time by monitoring run-to-failure bearing experiments on an FE8 bearing test rig. It is employed to detect blind faults in real time during ongoing experiments, resulting in the termination of the experiment to analyse the cause of fault initiation. Furthermore, historical data is utilised to quantify the lead time between the proposed method's detection and the final termination triggered by traditional condition monitoring methods. A comparative analysis of the physical bearing damage after the completion of tests demonstrates the capability of the proposed method to detect blind faults at an early stage. The results suggest that this approach identifies fault at an earlier damage stage as compared to traditional methods. It improves the conditions for studying bearing fatigue initiation by avoiding the interference of secondary damage or extensive spalling. In addition, the real-time blind fault detection capability demonstrates the practical application of the proposed framework in preventing catastrophic failures within high-value industrial assets.

1. INTRODUCTION

Early fault detection is crucial when monitoring critical mechanical components. It avoids unplanned downtime while providing sufficient lead time to plan maintenance strategies during operation. Vibration analysis is one of the most commonly used approaches for fault detection of rotating machinery (Peeters, Leclère, et al., 2019). One way to monitor these vibrations is through statistical indicators that characterise the statistical properties of the vibration signal over time. Examples include the root mean square, kurtosis, peak-to-peak, peak energy index, and crest factor (Peeters, Verstraeten, Nowé, & Helsen, 2019; Antoni et al., 2024; Večeř, Kreidl, & Šmíd, 2005). Alternatively, frequency domain indicators monitor specific characteristic frequencies introduced by the faults (Mohd Ghazali & Rahiman, 2021). Examples include the characteristic fault frequencies of bearings and gears, such as the ball pass frequency of the outer and inner race (Randall & Antoni, 2011), the gear mesh frequency, and its associated spectral sidebands (Sharma & Parey, 2016). Typically, domain experts in the field of mechanics are required to perform complex analyses on these indicators to monitor the health of mechanical components. Therefore, monitoring complex machines, such as wind turbines on a complete wind farm level in the real world becomes challenging. They contain many different components, including bearings, gears, and shafts, requiring specific monitoring techniques for fault detection. The number of health monitoring indicators tailored to specific components and fault types can therefore reach thousands. As a result, scaling the manual monitoring approach to a fleet of complex machines becomes impractical. Artificial intelligence can be used to automate the monitoring process by fusing numerous condition indica-

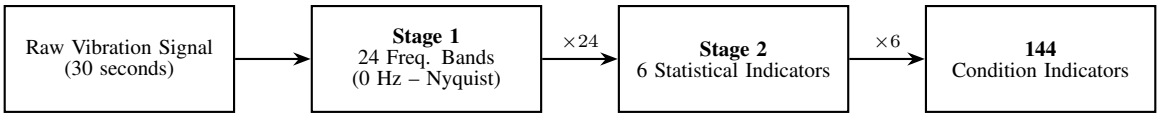


Figure 1. Overview of the signal processing pipeline to compute 144 condition indicators from the raw vibration signal.

tors into a single high-level health indicator (Jamil, Peeters, Verstraeten, & Helsen, 2025).

Machine learning methods are employed to automate the complex task of condition monitoring. The machine learning-based condition monitoring methods are divided into three categories (Cen, Yang, Liu, Xiong, & Chen, 2022): shallow machine learning (Saari, Strömbergsson, Lundberg, & Thomson, 2019), deep learning (Lei et al., 2020), and transfer learning (Jamil, Verstraeten, Nowé, Peeters, & Helsen, 2022) methods. The practical application of supervised learning in condition monitoring is often limited by the difficulty of obtaining balanced datasets that adequately represent the full spectrum of potential failure modes. Therefore, an unsupervised normal behaviour model, which requires only healthy data for training, presents a more suitable and scalable solution for real-time monitoring. A linear regression-based normal behaviour method derives a health indicator by fusing individual condition indicators labelled as healthy, warning, or faulty, utilising both SCADA and vibration data (Jamil et al., 2025). Similarly, a multivariate autoencoder is trained on time-domain statistical (Jamil, Jara Avila, Vratsinis, Peeters, & Helsen, 2023) and frequency-domain condition indicators (Jamil, Peeters, Verstraeten, & Helsen, 2023), integrated with SCADA operating conditions, to generate a single high-level health indicator for monitoring mechanical components.

This research presents an experimental study, using bearing fatigue testing on a test rig to validate the blind early fault detection capabilities of the multivariate autoencoder method. The approach is evaluated using two distinct frameworks: first, by detecting faults within an offline historical dataset, where experiments were originally terminated using traditional condition monitoring, and second, through real-time monitoring, where the proposed method is employed to trigger experiment termination. Finally, a comparative analysis of the physical damage to the bearings is conducted to demonstrate the ability of the proposed method in identifying faults at an incipient stage. The research focuses on blind early fault detection during real-time monitoring, Stammler et al. discusses the effectiveness of the proposed method in identifying faults at an early stage when compared with the classical condition monitoring techniques. Furthermore, the proposed framework fully automates the condition monitoring pipeline, eliminating the need for manual adjustment. In contrast, classical condition monitoring requires significant efforts and domain expertise to select, configure, and tune an appropriate fault detection method for a specific test case.

2. MULTIVARIATE AUTOENCODER

The research proposes a normal behaviour multivariate autoencoder for mechanical health monitoring, which fuses numerous traditional condition indicators into a single high-level health indicator. The high-level health indicator replaces the requirement to individually monitor every condition indicator and provides a simplified health indicator for operators to assist in planning efficient maintenance strategies. The condition indicators calculated from the vibration signal measured by an accelerometer are used as input to the model. The model is trained on healthy data, measured during the expected machine operations. It learns the normal behaviour of the machine from healthy data, and during monitoring, triggers fault alarms when it observes deviation from the learned normal behaviour. The proposed method is divided into three key steps: data measurement and vibration signal processing, multivariate deep learning, and fault alarm generation.

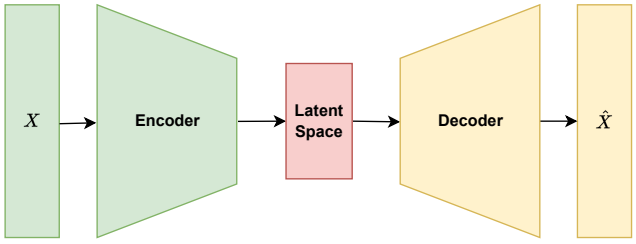


Figure 2. Overview of the autoencoder architecture, consisting of three main components: the encoder, latent space, and decoder. The model is trained to reconstruct an approximation $\hat{X}$ of the input X.

2.1. Data measurement and vibration signal processing

The raw vibration signal measured from an accelerometer is employed as the input to the method. Each vibration signal measurement is divided into 30-second intervals to calculate the condition indicators. The condition indicators are calculated in two stages, as illustrated in Figure 1. In the first stage, 24 frequency bands from 0 Hz to the Nyquist frequency are computed to increase the frequency-based sensitivity of the condition indicators (Antoni, 2007). In the second stage, six statistical indicators, root mean square, crest factor, kurtosis, number of 4-sigma outliers, time negentropy, and spectral negentropy, are calculated for every computed frequency band during the first stage (Peeters, Verstraeten, et al., 2019). The signal processing pipeline derives 144 condition indicators from a single raw vibration signal and serves

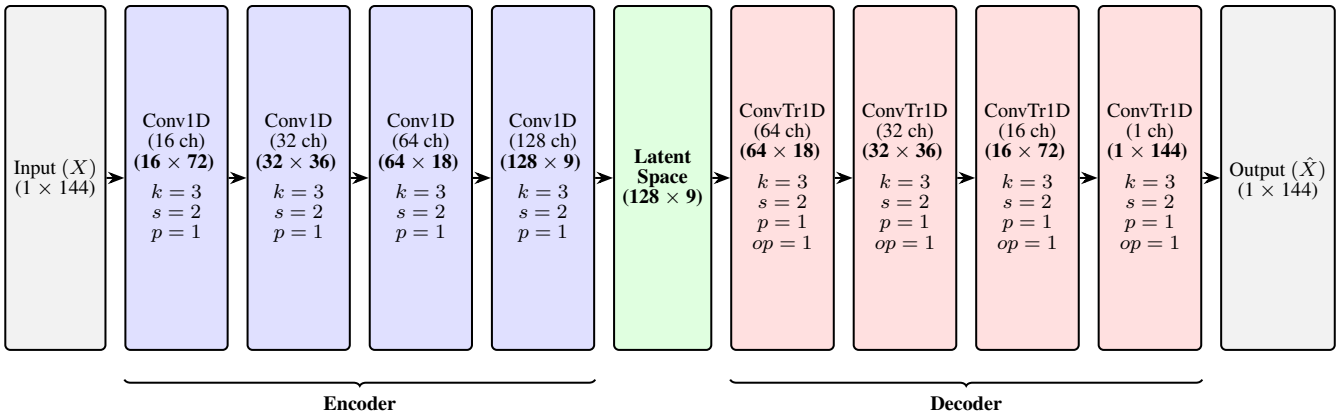


Figure 3. Complete architectural layout of the multivariate autoencoder, which employs 1D convolutional layers to transform a 144-indicator input vector into a 128×9 latent representation, increasing channel depth while reducing spatial dimensionality. The original signal is subsequently reconstructed through symmetrical transpose convolutional layers. The hyperparameters of 1D convolutional layers are mentioned as: kernel k, stride s, padding p, and output padding op.

as the input to the multivariate autoencoder model. While 144 condition indicators were selected for this study, the method allows the inclusion of further time- and frequency-domain indicators if required. The main consideration is that these selected indicators capture, to some extent, the degree of non-stationarity and non-Gaussianity in the measured vibration signals (Antoni et al., 2024), as these characteristics are commonly associated with fault-related behaviour.

2.2. Multivariate deep learning

Deep learning models consist of numerous interconnected computational units (neurons) arranged in multiple processing layers, enabling them to learn complex data representations across different levels of abstraction (LeCun, Bengio, & Hinton, 2015). The proposed method employs a deep learning–based multivariate autoencoder to combine multiple condition indicators into a single high-level health indicator. Autoencoders, as illustrated in Figure 2, are unsupervised neural networks designed to encode the input into a lower-dimensional latent space representation, and then decode an approximation of the input from the compressed latent space. The latent space represents the important features required for reconstruction. Autoencoders are trained to minimise the reconstruction error between the input and output, and are used for unsupervised tasks, such as anomaly detection, feature extraction, and data generation (Mienye & Swart, 2025). The proposed normal-behaviour multivariate autoencoder model is trained on data collected during the machine's normal operation. This healthy data can be identified by considering the mean time between failures (MTBF) and the mean time to failure (MTTF), while excluding the machine's initial run-in period. Alternatively, experts can identify the healthy operating window by observing operational trends. Selecting appropriate training data from a healthy operation is critical; if data from initial fault stages are included in the training set, the model learns this as normal behaviour, significantly delaying alarm triggers during fault progression. However, the model demonstrates robustness against small proportions of fault data and outliers, as validated by the inclusion of the run-in period in the experimental datasets. The healthy dataset D is a time series comprising M observations:

$$D = \{t_i | i = 1, ..., M\} \tag{1}$$

where each timestamp t_i is a set of N condition indicators calculated from a single vibration signal. Each timestamp t_i is represented as a vector X_i that contains N condition indicators, as defined in Equation 2.

$$X_i = \{x_1, x_2, x_3, ...x_N\} \tag{2}$$

The multivariate autoencoder is defined as a composite function consisting of an encoder f and a decoder g, trained jointly to optimise the encoder parameters ϕ and decoder parameters α.

$$\hat{X}_i = g_\alpha(f_\phi(X_i)) \tag{3}$$

The $\hat{X}_i$ is the approximation of the input X_i:

$$\hat{X}_i = \{\hat{x_1}, \hat{x_2}, \hat{x_3}, ...\hat{x_N}\} \tag{4}$$

The loss function $\mathcal{L}$ to calculate the reconstruction error between the input X_i and the reconstructed output $\hat{X}_i$ is defined as follows:

$$\mathcal{L} = \frac{1}{N}\sum_{j=1}^{N}(x_j - \hat{x}_j)^2 \tag{5}$$

N represents the number of condition indicators, x_j and $\hat{x}_j$ represent a single corresponding condition indicator value from the input Equation 2 and output Equation 4 vectors, respectively.

The normal behaviour multivariate autoencoder, trained ex-

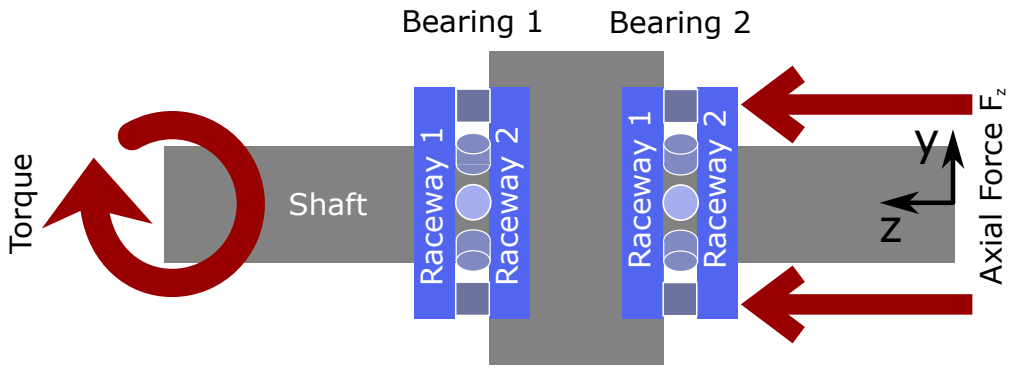


Figure 4. Layout of the FE8 bearing test rig on which the fatigue testing of bearings were conducted

clusively on healthy data, develops a bias toward accurately reconstructing healthy behaviour patterns with minimal error. In contrast, it struggles to reconstruct faulty behaviour, resulting in a significantly higher reconstruction error. Therefore, the reconstruction error is used to derive a high-level health indicator, representing a single point in the multidimensional condition indicator space. A standalone model is trained for each test case using only its initial healthy operational data to account for variations in vibration signatures across different experiments. This strategy mitigates the influence of change in the vibration signature and ensures high sensitivity to deviation from the individual experiment's normal behaviour.

The complete architecture of the proposed multivariate autoencoder is illustrated in Figure 3. The 1D convolutional layers enable the model to capture hierarchical relationships between the 24 frequency bands and 6 statistical metrics. The architecture follows a symmetrical encoder-decoder structure. The encoder compresses the sequence length of the input from 144 to 9 through strided convolutions while expanding the channel depth from 1 to 128 to extract increasingly abstract features. The decoder then reconstructs the signal by mirroring the encoder layers using transposed convolutions.

2.3. Fault alarm generation

The high-level health indicator provides the health monitoring trend for machine's components. It is defined as the z-score value computed from the reconstruction error. The z-score is the relationship of a datapoint with the mean measured in terms of standard deviation, as defined in Equation 6.

$$z_i = \frac{\mathcal{L}_i - \mu}{\sigma} \tag{6}$$

Here, μ and σ represent the mean and standard deviation derived from the reconstruction error obtained from the healthy dataset D after the completion of training, as defined in Equation 7 and Equation 8, respectively.

$$\mu = \frac{1}{M}\sum_{i=1}^{M}\mathcal{L}_i \tag{7}$$

$$\sigma = \sqrt{\frac{1}{M}\sum_{i=1}^{M}(\mathcal{L}_i - \mu)^2} \tag{8}$$

During monitoring, condition indicators are computed from each raw vibration signal measurement and provided as input to the trained model to obtain the reconstruction error. The high-level health indicator is then calculated as a z-score using the μ and σ derived from the healthy training data, as defined in Equation 6. Fault alarms are generated when the z-score deviates significantly from the learned normal behaviour. Two threshold levels are used to classify the high-level health indicator into healthy, warning, and alarm states. These thresholds are set at z-score values of 5 and 10; values below 5 are labelled as healthy, values between 5 and 10 indicate a warning state, and values exceeding 10 are classified as an alarm state. Faults manifest at a significant distance from learned normal behaviour; therefore, the threshold levels are adjusted to minimise false alarms. During real-time monitoring, the indicator trend is continuously tracked, and experts are notified to terminate the test once the alarm threshold is surpassed.

3. EXPERIMENTS

The test rig used for conducting the experiments from which the data is collected is an FE8 bearing test rig manufactured by Christoph Assmann Engineering & Industrial Services with a layout used for testing as shown in Figure 4. It is used to conduct fatigue testing of the 81212-TV axial thrust roller bearings, manufactured by Schaeffler (Schaeffler, 2025), until failure. The properties of the bearings are listed in Table 2. The test conditions used can be seen in Table 1.

For detecting failure and stopping the test, the test rig in-

Table 1. Overview of the bearing test conditions.

Test ID	**Force** F_z (kN)	**Speed** ω (rpm)	**Temp.** T (°C)	**Stress** σ_{Hertz} (MPa)	**Oil used**
T1, T3–T5	85	250	100	1800	Mobil SHC Gear 320 WT
T2	95	250	100	1900	Mobil SHC Gear 320 WT

Table 2. Dimensions and properties of the 81212 axial thrust roller bearing. Except for the rolling element dimensions and the number of rollers, all data were obtained from (Schaeffler, 2025). The rolling element dimensions and roller count were determined by measurement.

Property	**Symbol**	**Value**	**Unit**
Bore diameter	d_i	60	mm
Outer diameter	d_o	95	mm
Width	B	26	mm
Number of rollers	z	19	–
Rolling element diameter	D_W	11	mm
Rolling element length	L_W	11	mm
Effective rolling element length	L_{we}	9	mm
Static load capacity	C_0	480,000	N
Dynamic load capacity	C_1	173,000	N

cludes an accelerometer Acida AC101.51 which monitors the acceleration amplitude. It is generally seen that after a run-in period of 12-24 hours of starting the test, the acceleration detected by the sensor decreases to 1.2 g below the initially found acceleration value after which the tests tend to have a steady acceleration amplitude till the initiation of failure. The stopping criteria used for detecting bearing failure, determined using previous bearing testing to ensure visible damage of the bearing after stopping the test, is an increase of 1 g in the acceleration amplitude above this steady acceleration amplitude.

The vibration spectrum data used for the analysis was recorded using a 356A03 piezoelectric accelerometer from PCB Piezoelectronics (*PCB Piezotronics*, 2025) installed on the housing of the test rig. This accelerometer was used in conjunction with a Simcenter SCADA lab system from Siemens Digital Industry Software for Tests T1, T2 and T3. For the case of tests T4 and T5, a custom-built condition monitoring system from VUB was used. This custom condition monitoring system is deployed by VUB across multiple offshore wind turbines for monitoring drivetrain health. For tests T3 and T5, the automatic stopping criteria remained consistent with previous experiments; however, the high-level health indicator trend was also monitored in real-time, allowing the tests to be manually terminated upon the detection of a fault. Data from test T4 is not used in further analysis of the results due to issues with data collection during the testing; the data was not stored at the beginning of the experiment. The raw vibration data used in this research are publicly available at the DTU data library (Sudhakaran & Jamil, 2026). The dataset comprises five complete bearing run-to-failure experiments, providing a robust basis for validating prognostics and health management (PHM) methods and estimating remaining useful life (RUL).

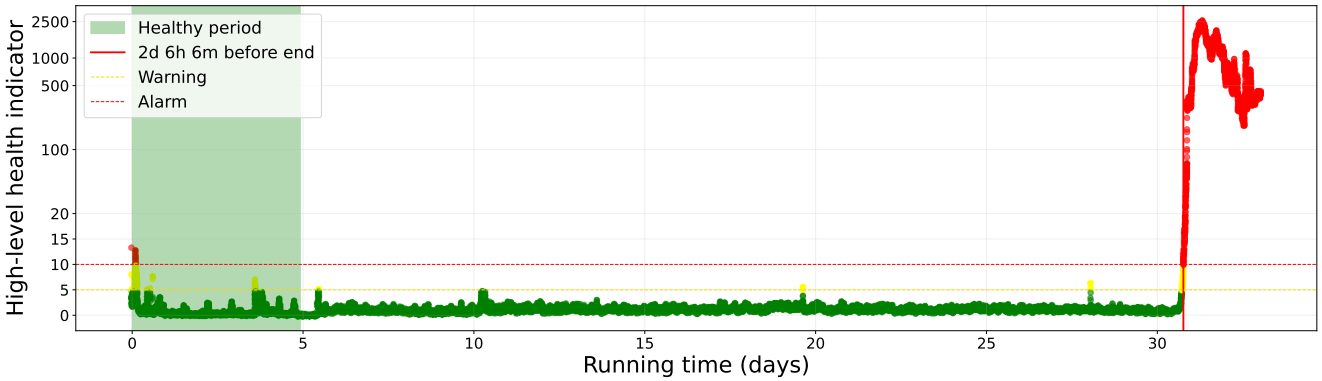


Figure 5. T1 temporal evolution of the high-level health indicator demonstrating the lead time achieved on historical data. The trend exhibits rapid progression due to critical damage, providing a lead time of 2 days, 6 hours, and 6 minutes prior to the traditional condition monitoring method triggering the system shutdown.

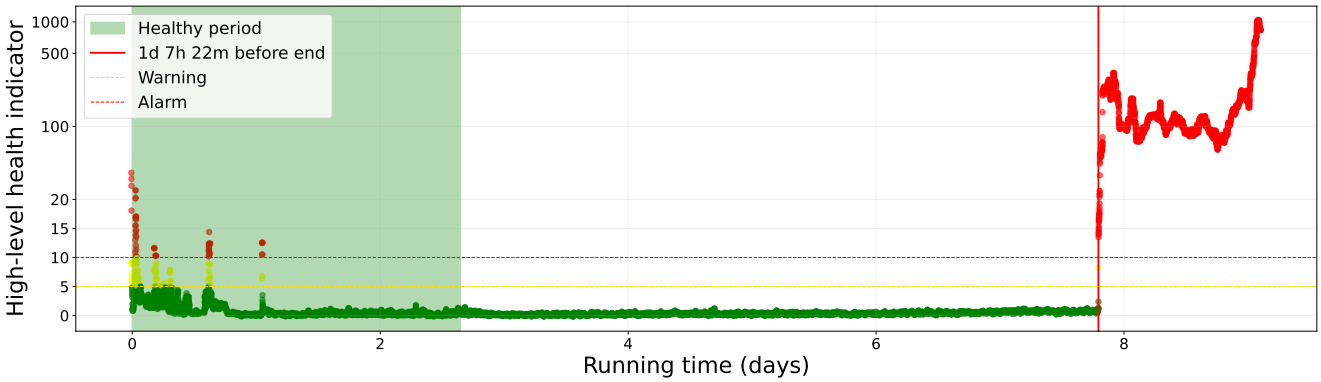


Figure 6. T2 temporal evolution of the high-level health indicator demonstrating the lead time achieved on historical data. The trend exhibits rapid progression due to critical damage, providing a lead time of 1 day, 7 hours, and 22 minutes prior to the traditional condition monitoring method triggering the system shutdown.

4. RESULTS

The proposed method is validated in real-time on an experimental test rig setup to demonstrate its blind early fault detection capability. The study follows a two-stage validation process. First, the method is applied retrospectively to recorded datasets, where a traditional condition monitoring approach was used to terminate the test. This allows for a direct comparison to quantify the lead time provided by the proposed method. Second, the method is deployed for live monitoring of ongoing tests, where the generated health indicator trend is monitored to perform a manual termination upon fault detection. Finally, the bearing physical conditions from both scenarios are analysed to confirm the effectiveness of the proposed method in identifying incipient faults.

4.1. Offline lead-time analysis

The proposed method is validated using two historical test datasets to establish a performance baseline before real-time deployment. These results are quantified in terms of lead time relative to the traditional condition monitoring methods that originally triggered the test shutdowns. Table 3 provides an overview of the experiment duration and lead time for each test. To accurately reflect the component's degradation, the analysis distinguishes between total duration and running time. The total duration is the entire experimental period, including standby periods and maintenance breaks. In contrast, running time represents the actual operational runtime, excluding every non-active period. The lead time is determined as the interval between the initial point where the high-level alarm trend persistently exceeds the alarm threshold and the final termination of the test. The proposed method identifies

Table 3. Lead-time quantification across offline test datasets. d, h, and m denote days, hours, and minutes, respectively.

Test ID	Total duration	Running time	Lead Time
T1	41d, 40h, 31m	33d, 0h, 24m	2d, 6h, 6m
T2	13d, 13h, 37m	9d, 2h, 24m	1d, 7h, 22m

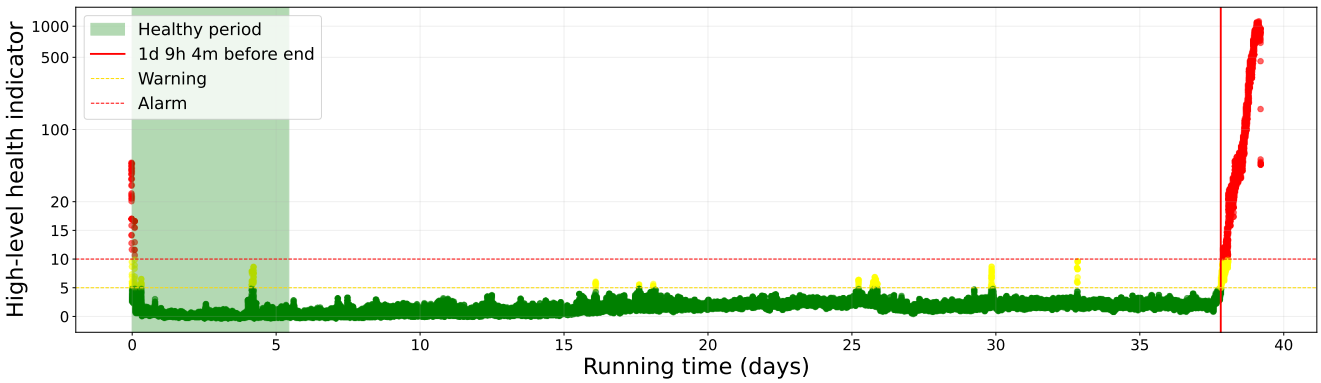


Figure 7. Real-time monitoring of the health indicator trend for T3; the test is terminated manually once the health trend surpasses the alarm threshold.

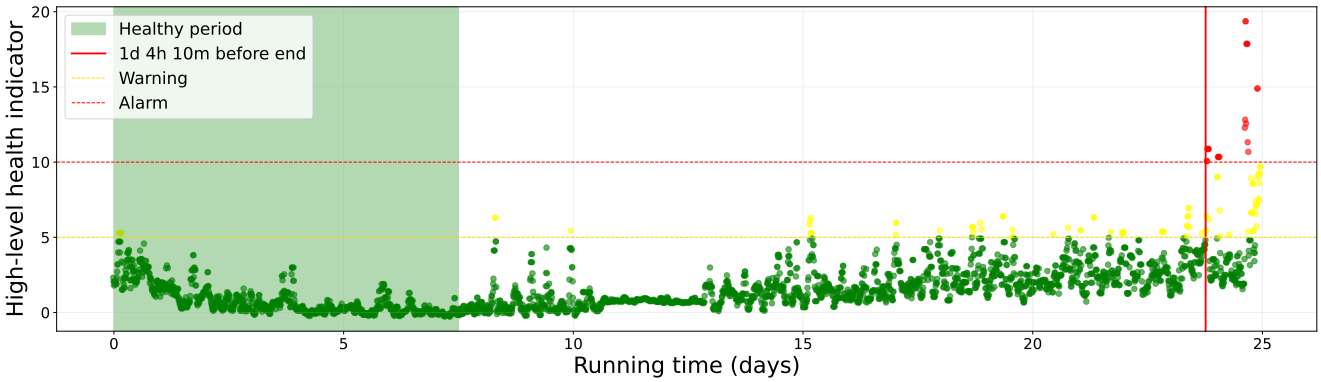


Figure 8. Real-time monitoring of the health indicator trend for T5; the test is terminated manually once the health trend surpasses the alarm threshold.

a critical fault with a lead time of 2 days, 6 hours, and 6 minutes in T1, as illustrated in Figure 5. Similarly, in T2, the high-level health indicator trend captures the transition to a faulty state 1 day, 7 hours, and 22 minutes prior to the test termination as depicted in Figure 6. In both cases, the critical damage causes an abrupt change in machine vibrational behaviour. However, the proposed method successfully captures the deviation from the learned normal behaviour, offering sufficient warning before the traditional condition monitoring method triggers termination. These results demonstrate the model's robustness and its capability to provide a significant early warning window, even in the presence of accelerated failure modes where the margin for intervention is narrow.

4.2. Real-time monitoring

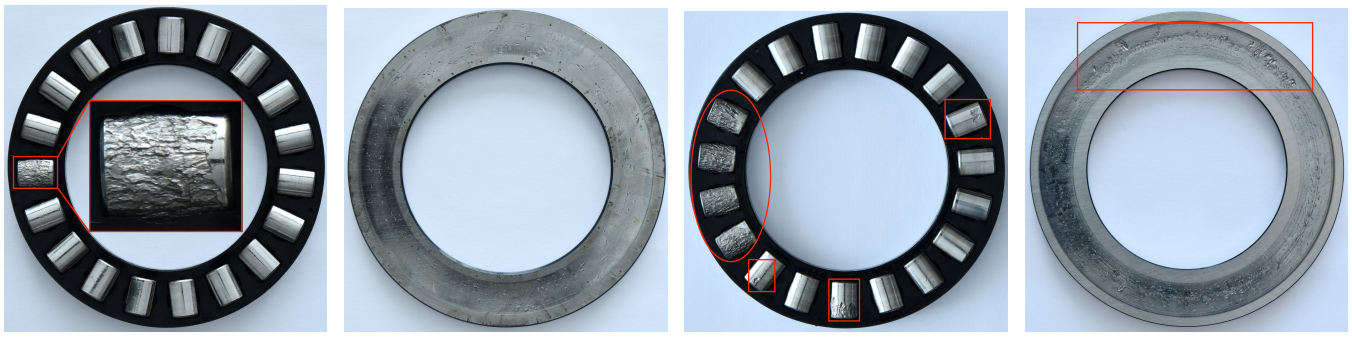
Lead-time analysis performed on offline historical data establishes a performance baseline for real-time deployment, enabling informed experimental termination upon early detection of a fault. The proposed method is subsequently deployed to monitor two additional test cases in real time to further validate its early fault detection capability. During these tests, the high-level health indicator trend is continuously observed, and the experimental setup is terminated only after the presence of a fault is confirmed through a persistent fault trend. Table 4 provides an overview of the total duration and running time of the experiment before the termination for real-time monitoring tests. To assess the effectiveness of the approach, the resulting physical damage of the bearings is compared with the damage observed in the offline tests, where shutdown is triggered using traditional condition monitoring techniques. This comparison demonstrates that the proposed method detects faults at an early stage during incipient damage, providing sufficient lead time to plan predictive maintenance strategies in real-world applications. Figures 7 and 8 illustrate the high-level health indicator trends for the real-time monitoring tests T3 and T5, respectively, where the test is terminated after the indicator trend exceeds the alarm threshold level. The monitoring system does not have direct control of the test rig; instead, the shutdown is performed manually following the observation of a persistent fault trend

Table 4. The real-time monitoring experimental datasets overview. d, h, and m denote days, hours, and minutes, respectively.

Test ID	Total duration	Running time
T3	40d, 20h, 8m	39d, 04h, 33m
T5	48d, 22h, 50m	24, 22h, 30m

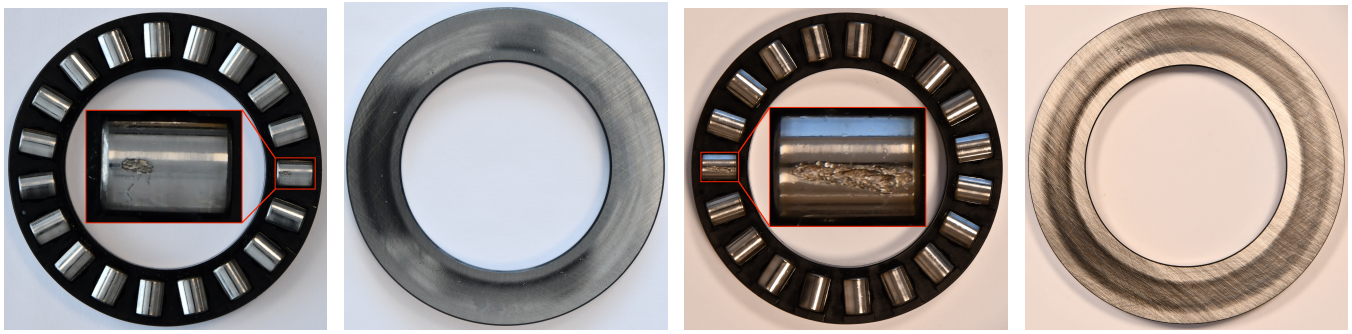
4.3. Comparative Analysis of Physical Bearing Damage

A comparative analysis is performed on the physical state of the bearings after the completion of tests to demonstrate the effectiveness of the proposed method in detecting blind early faults. The observed bearing damage serves as a degradation metric to compare real-time monitoring test cases, where the proposed method triggers the termination, against offline test cases stopped by employing traditional condition monitoring methods. The overview of observed bearing damages across all four tests is provided in Table 5. The bearing condition following the completion of tests T1 and T2 is depicted in Figure 9. As discussed in Section 4.1, traditional monitoring methods detect faults only at a severe stage before triggering termination, resulting in significant physical damage. In both cases, damage is visible on the raceways and the rolling elements. Specifically, the T1 bearing exhibits severe spalling on a single roller. The T2 bearing displays even more extensive degradation, with severe spalling on four rollers, spalling on three rollers, and a longitudinal spall along the raceway. In contrast, Figure 10 illustrates the bearing condition for tests T3 and T5, which are monitored in real-time using the proposed multivariate autoencoder. The test rig is stopped upon the confirmation of fault presence via the high-



(a) T1 bearing rollers and cage (b) T1 bearing raceway (c) T2 bearing rollers and cage (d) T2 bearing raceway

Figure 9. Severe bearing degradation is observed in tests terminated by traditional condition monitoring methods. The images illustrate extensive surface fatigue, characterised by large-scale spalling and significant material loss.



(a) T3 Bearing rollers and cage (b) T3 bearing raceway (c) T5 bearing rollers and cage (d) T5 bearing raceway

Figure 10. Incipient bearing damage is observed in real-time monitoring test cases terminated by employing the proposed multivariate autoencoder method. The images depict localised micro-pitting and early-stage surface fatigue, demonstrating the model's ability to identify faults at an initial phase of degradation.

level health indicator, ensuring that physical damage remains minimal. Consequently, degradation in both T3 and T5 is confined to a single roller, with no visible damage observed on the raceways. The comparison of the physical condition of the bearings suggests that the proposed multivariate autoencoder can identify incipient faults at an early stage during real-time monitoring.

Table 5. Bearing damages observed after the completion of tests.

Test ID	Damage on bearing
T1	Bearing 2: Advanced spalling on one roller.
T2	Bearing 1: Advanced spalling on multiple rollers; a longitudinal spall along the raceway.
T3	Bearing 2: Small damage marks on one roller.
T5	Bearing 2: Small damage marks on one roller.

5. DISCUSSION

This research presents an experimental study conducted on an FE8 bearing test rig to validate the early blind fault detection capabilities of a multivariate autoencoder. The proposed method is a physics-based deep learning normal behaviour model which fuses numerous traditional condition indicators into a single high-level health indicator. These condition indicators are derived from raw vibration signals measured by an accelerometer and processed using advanced signal processing techniques. By aggregating these various condition indicators, the model provides a comprehensive real-time overview of the machine's health status. To ensure the robustness of the proposed method in real-life scenarios, the research follows a two-step validation process that utilises available historical data before deployment in a live experimental setting. In the first step, historical datasets from tests T1 and T2 are analysed to validate the model's sensitivity in detecting faults at an early stage by determining the potential lead time. This lead time is defined as the interval between the model's initial detection and the eventual termination triggered by traditional monitoring methods. The second step involves real-time monitoring, where the proposed method is used to monitor the health status of the ongoing bearing tests T3 and T5. At this stage, the model's predictions are employed to actively monitor the ongoing experiment; once a fault is indicated by the model, a manual termination is

triggered. The results suggest that the proposed method detects faults in the incipient stage before progressing to severe physical damage. A comparative analysis is performed to assess the extent of bearing damage following the completion of the tests. Table 5 provides an overview of the degradation observed across all four tests. As depicted in Figure 9, the bearings observed severe damage when tests T1 and T2 were terminated using traditional condition monitoring methods. In contrast, the bearings show minimal damage in tests T3 and T5, where the proposed method is employed for real-time monitoring, as shown in Figure 10. The results suggest that the proposed method detects faults at an early stage before they progress to severe physical damage. Preventing the spalling damage from progressing also helps to better study the initiation of the bearing fatigue damage. As the prevalent source of bearing fatigue involves the formation of cracks forming below the bearing surface and travelling to the surface to cause spalling, extensive damage on the surface could lead to a lower probability of finding these initial cracks as they would be removed in further operation.

The proposed method simplifies the health monitoring process. It provides a high-level overview of the machine components' health. Real-time monitoring is critical in preventing catastrophic failures in high-value industrial assets. Fusing condition indicators into a single high-level health indicator involves a trade-off. This can mask the underlying condition indicator level trends that experts use for detailed root cause analysis. Future research will focus on integrating explainable artificial intelligence techniques. These will identify and rank specific condition indicators that contribute to the detected fault. As a result, the method will offer dual benefits. It will provide transparent, in-depth diagnostic analysis for domain experts and a simplified overview for non-expert operators.

6. CONCLUSION

The research presented a normal behaviour multivariate autoencoder method for the early fault detection of mechanical components. The proposed method simplified the condition monitoring process by fusing numerous condition indicators into a single high-level health indicator, providing a comprehensive overview of component health. The normal behaviour model was trained on healthy data to learn the normal behaviour of the machine and detect faults by identifying deviations from this learned behaviour. The method was validated through real-time bearing run-to-failure experiments, where it assisted in experiment termination upon fault detection. Furthermore, the model was compared to traditional condition monitoring methods using historical data to quantify the lead time gained. A comparative study of the bearing damage following the completion of the tests demonstrated that the proposed method identified faults at a more incipient stage as compared to traditional methods, which detected them only after the occurrence of more severe damage. Consequently, the ability to detect initial spalling helps stop further fatigue progression and thereby preserves the early stage damage provides an advantage for studying the initiation of fatigue.

7. DATA AVAILABILITY

The raw vibration data used in this research are derived from five bearing run-to-failure experiments. These datasets are publicly available through the DTU Data Library at the following persistent link:
https://doi.org/10.11583/DTU.31851214.

ACKNOWLEDGMENT

The authors would like to acknowledge FWO (Fonds Wetenschappelijk Onderzoek) for their support through the SB grants of Faras Jamil (#1S63123N) and the long stay abroad of Faras Jamil (#V405425N). Furthermore, the authors would also like to acknowledge FOD Economy for their support through the Energy Transition Fund project SWIFT. They also acknowledge VLAIO for their support in the context of the Blue Cluster, for the ICON project Supersized 5.0 and the cSBO project CORE. The authors would also like to acknowledge the financial support from the Danish Energy Technology Development and Demonstration Program (EUDP) through the projects "MODA: Modularisation and Data-driven Design of Test Infrastructure for the Wind Industry" (journal no. 640242-534569) and "BOREAS: Improving Blade Bearing Reliability Through Big rotOR blade bEAring test Strategy" (journal no. 640241-521726).

REFERENCES

Antoni, J. (2007). Fast computation of the kurtogram for the detection of transient faults. *Mechanical Systems and Signal Processing*, *21*(1), 108–124. doi: 10.1016/j.ymssp.2005.12.002

Antoni, J., Kestel, K., Peeters, C., Leclère, Q., Girardin, F., Ooijevaar, T., & Helsen, J. (2024). On the design of optimal health indicators for early fault detection and their statistical thresholds. *Mechanical Systems and Signal Processing*, *218*, 111518.

Cen, J., Yang, Z., Liu, X., Xiong, J., & Chen, H. (2022, Oct 01). A review of data-driven machinery fault diagnosis using machine learning algorithms. *Journal of Vibration Engineering & Technologies*, *10*(7), 2481-2507. doi: 10.1007/s42417-022-00498-9

Jamil, F., Jara Avila, F., Vratsinis, K., Peeters, C., & Helsen, J. (2023, 06). Wind turbine drivetrain fault detection using multi-variate deep learning combined with signal processing. In *Turbo expo: Power for land, sea, and air* (Vol. Volume 14: Wind Energy, p. V014T37A003). Retrieved from

https://doi.org/10.1115/GT2023-101689 doi: 10.1115/GT2023-101689

Jamil, F., Peeters, C., Verstraeten, T., & Helsen, J. (2023, July). Wind turbine drivetrain fault detection using physics-informed multivariate deep learning. In *Surveillance, Vibrations, Shock and Noise*. Toulouse, France. Retrieved from https://hal.science/hal-04166103

Jamil, F., Peeters, C., Verstraeten, T., & Helsen, J. (2025). Leveraging signal processing and machine learning for automated fault detection in wind turbine drivetrains. *Wind Energy Science*, *10*(9), 1963–1978. doi: https://doi.org/10.5194/wes-10-1963-2025

Jamil, F., Verstraeten, T., Nowé, A., Peeters, C., & Helsen, J. (2022). A deep boosted transfer learning method for wind turbine gearbox fault detection. *Renewable Energy*, *197*, 331-341. doi: https://doi.org/10.1016/j.renene.2022.07.117

LeCun, Y., Bengio, Y., & Hinton, G. (2015, May 01). Deep learning. *Nature*, *521*(7553), 436-444. doi: 10.1038/nature14539

Lei, Y., Yang, B., Jiang, X., Jia, F., Li, N., & Nandi, A. K. (2020). Applications of machine learning to machine fault diagnosis: A review and roadmap. *Mechanical Systems and Signal Processing*, *138*, 106587. doi: https://doi.org/10.1016/j.ymssp.2019.106587

Mienye, I. D., & Swart, T. G. (2025, October). Deep Autoencoder Neural Networks: A Comprehensive Review and New Perspectives. *Archives of Computational Methods in Engineering*, *32*(7), 3981–4000. Retrieved 2026-02-20, from 10.1007/s11831-025-10260-5 doi: 10.1007/s11831-025-10260-5

Mohd Ghazali, M. H., & Rahiman, W. (2021). Vibration analysis for machine monitoring and diagnosis: a systematic review. *Shock and Vibration*, *2021*(1), 9469318. doi: 10.1155/2021/9469318

Pcb piezotronics. (2025). Retrieved from https://www.pcb.com/products?m=356a03

Peeters, C., Leclère, Q., Antoni, J., Lindahl, P., Donnal, J., Leeb, S., & Helsen, J. (2019). Review and comparison of tacholess instantaneous speed estimation methods on experimental vibration data. *Mechanical Systems and Signal Processing*, *129*, 407-436. doi: 10.1016/j.ymssp.2019.02.031

Peeters, C., Verstraeten, T., Nowé, A., & Helsen, J. (2019, 06). Wind turbine planetary gear fault identification using statistical condition indicators and machine learning. In *International conference on offshore mechanics and arctic engineering* (Vol. Volume 10: Ocean Renewable Energy). doi: 10.1115/OMAE2019-96713

Randall, R. B., & Antoni, J. (2011). Rolling element bearing diagnostics—a tutorial. *Mechanical Systems and Signal Processing*, *25*(2), 485-520. doi: 10.1016/j.ymssp.2010.07.017

Saari, J., Strömbergsson, D., Lundberg, J., & Thomson, A. (2019). Detection and identification of windmill bearing faults using a one-class support vector machine (svm). *Measurement*, *137*, 287-301. doi: 10.1016/j.measurement.2019.01.020

Schaeffler. (2025). *81212-tv axial cylindrical roller bearings*. Retrieved 2025-03-25, from https://medias.schaeffler.dk/

Sharma, V., & Parey, A. (2016). A review of gear fault diagnosis using various condition indicators. *Procedia Engineering*, *144*, 253-263. (International Conference on Vibration Problems 2015) doi: 10.1016/j.proeng.2016.05.131

Stammler, M., Jamil, F., Liu, X., Matthys, J. J., Sudhakaran, N., Peeters, C., ... Helsen, J. (2026). Ai enhanced fault indicators vs. classical bearing monitoring – example results of bearing tests and transferability to wind turbines. *Wind Energy Science Discussions*, *2026*, 1–27. Retrieved from https://wes.copernicus.org/preprints/wes-2026- doi: 10.5194/wes-2026-81

Sudhakaran, N., & Jamil, F. (2026, 4). Vibration Data from 81212 Bearing fatigue testing on FE8 Test Rig. Retrieved from https://doi.org/10.11583/DTU.31851214 doi: 10.11583/DTU.31851214.v1

Večeř, P., Kreidl, M., & Šmíd, R. (2005, Jan.). Condition indicators for gearbox condition monitoring systems. *Acta Polytechnica*, *45*(6). doi: 10.14311/782