https://doi.org/10.5194/wes-2026-83
Preprint. Discussion started: 6 July 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

SCADA-free wind turbine drivetrain health monitoring using a physics-informed multivariate autoencoder

Faras Jamil[1,2], Cédric Peeters[1,3], and Jan Helsen[1,3]

[1]Acoustics & Vibration Research Group / OWI-Lab, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussel, Belgium
[2]Artificial Intelligence Lab, Vrije Universiteit Brussel, Pleinlaan 9, 1050 Brussel, Belgium
[3]Flanders Make@VUB, Flanders Make, Oude Diestersebaan 133, 3920 Lommel, Belgium

Correspondence: Faras Jamil (faras.jamil@vub.be)

Abstract. Effective condition monitoring is critical for preventing costly machine failures. Vibration analysis is one of the most widely adopted condition monitoring approaches. It enables early fault detection by capturing subtle dynamic changes caused by misalignment, imbalance, and bearing wear. Traditional techniques rely on signal processing in the time and frequency domains, and experts manually track individual condition indicators to identify fault trends. The manual tracking of condition indicators becomes impossible when monitoring large fleets of complex machines, such as wind turbines. This research proposes a novel approach to consolidate multiple condition indicators into a single high-level health indicator to simplify the monitoring process. A physics-informed multivariate autoencoder models the machine's normal behaviour using vibration-based condition indicators computed from the vibration signal measured during healthy operation. The non-linear model incorporates operating conditions from vibration condition indicators, without using SCADA as input. It identifies faults by detecting deviations from the established normal behaviour. The proposed method is validated on NASA's Intelligent Maintenance Systems bearing dataset and a multi-year offshore wind farm dataset with confirmed fault cases. Validation on a wind turbine drivetrain dataset demonstrated that the proposed method detects 100% (19/19) of labelled fault cases. The model achieves 97% balanced accuracy, and threshold optimisation further reduces false positives to 1 case out of 91 healthy cases, while maintaining high diagnostic sensitivity with only a single false negative (missed fault alarm). Results demonstrate that the proposed method reliably accommodates diverse condition indicators, effectively detects faults, and reduces the time required for condition monitoring analysis.

1 Introduction

The impact of global warming has led to a significant growth in renewable energy sources. The rising interest has directed research to produce cost-efficient renewable energy. Wind energy is considered a sustainable renewable energy source, as wind is a ubiquitously available resource for energy production. However, operating and maintenance costs are a substantial challenge for the wind energy industry. The continuous movement of the wind turbine's drivetrain rotatory components increases the risk of faults. As a result, these rotating components require constant monitoring.

Vibration analysis is the most commonly used method for monitoring the health of rotating components Peeters et al. (2019b), as the vibration signal is sensitive enough to detect even minor damage. However, vibration signals are susceptible to

https://doi.org/10.5194/wes-2026-83
Preprint. Discussion started: 6 July 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

interference and noise from other mechanical, electrical, and environmental sources. Different mechanical faults, ranging from bearing wear to gear damage, manifest unique vibration signatures. To identify these potential faults, condition monitoring techniques are employed to highlight associated patterns with them. These techniques involve signal filtering to enhance the signal-to-noise ratio and the computation of specialised signal processing indicators to monitor specific faults.

While it is possible to detect faults using signal processing indicators, each indicator is typically sensitive to specific types of faults; however, no single indicator can detect all kinds of faults. Statistical indicators track changes in the statistical properties of vibration signals over time. Frequency-domain indicators monitor the appearance of new frequency components introduced by faults. Time-domain statistical indicators detect faults caused by loose parts by observing changes in the amplitude of the vibration signal. Impact faults, such as those caused by a broken gear tooth or a bearing cage failure, are detectable in the time domain. Cavitation in pumps and compressors, introduced by pressure fluctuations, can also be observed through changes in the time domain. In some cases, bearing faults can cause changes in statistical properties, such as increased kurtosis. Frequency-domain indicators can detect faults like gear tooth damage, which introduces new frequency components. Frequency-domain indicators can also detect misalignment patterns in the components. Bearing faults, including ball bearing, roller bearing, inner race, and outer race, produce characteristic frequencies visible in frequency spectra. Rotor imbalance creates a dominant frequency component which is easily detectable. Considering the diverse types of faults that can affect rotating components, relying on a combination of signal-processing indicators is essential for comprehensive monitoring. Various time-domain and frequency-domain signal processing indicators are used to monitor fault trends.

The most commonly used foundational time-domain statistical indicators include root mean square (RMS), kurtosis, moors kurtosis, peak-to-peak, peak energy index (PEI), and crest factor Večeř et al. (2005); Peeters et al. (2019c); Mohd Ghazali and Rahiman (2021); Antoni et al. (2024b). RMS represents the signal energy and amplitude in the time domain Večeř et al. (2005), with a higher RMS value indicating a greater vibration level due to a progressive fault Igba et al. (2016). Kurtosis measures the tailedness in the vibration signal's probability distribution, where high kurtosis suggests the presence of extreme values caused by, e.g. bearing damage or loose parts Antoni and Randall (2006). Moors kurtosis, similar to standard kurtosis but based on quantiles Moors (1988), is less sensitive to outliers. Peak-to-peak is the measure of spread in the signal, determined by calculating the difference between the maximum and minimum amplitude values of the signal Zhu et al. (2014). It detects faults caused by impacts or transient events resulting in substantial amplitude variations. The PEI measures the energy in the largest peaks in a vibration signal Peeters et al. (2019c). A high PEI indicates energy is concentrated in these components, often due to sharp transients from bearing and gear faults, while a low PEI suggests more evenly distributed energy across the frequency spectrum. The crest factor is the ratio of the maximum peak value to the RMS value Mathew and Alfredson (1984), indicating the signal's spikiness caused by transient events or non-linear distortions. Frequency domain indicators monitor the amplitude of the vibration signal in the frequency spectrum, which allows the detection and diagnosis of mechanical faults. Frequency domain analysis allows for the observation of several signal characteristics that may not be visible in the time domain Mohd Ghazali and Rahiman (2021). Rotating components like gears, bearings, and shafts have associated characteristic frequencies, and tracking them provides critical insights about potential faults. For example, the ball pass frequency of the outer race (BPFO) and ball pass frequency of the inner race (BPFI) can indicate faults, such as micropitting, in the outer and inner

https://doi.org/10.5194/wes-2026-83
Preprint. Discussion started: 6 July 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

races of bearings. Similarly, the ball spin frequency (BSF) and fundamental train frequency (FTF) track faults in the bearing's rolling elements and cage Randall and Antoni (2011). The gear damage is detectable by observing characteristic frequencies such as the mesh frequency, gear rotational frequency, and modulation frequencies Sharma and Parey (2016). Tracking the harmonics in the autopower and envelope spectra monitors the health of gears and shafts, identifying potential degradation and damage. The health of gears and shafts can be monitored by observing the autopower and envelope spectra harmonics. Early detection of these faults is crucial to preventing equipment failures and ensuring the continued reliability and efficiency of the machinery.

Faults in their early stages induce changes in the vibration signal, which are challenging to detect. Experts are required to monitor the signal processing indicators to identify subtle changes. Observing multiple indicators for each component in a complex machine, such as a wind turbine operating under non-stationary conditions, presents a significant challenge. In a wind farm with numerous wind turbines, the number of indicators can reach into the thousands, which makes regular monitoring impractical. In such cases, machine learning algorithms identify the changes in signal-processing indicators related to faults. Trend analysis of the indicator changes over time helps detect signs of early-stage degradation. However, a balanced distribution from all classes is required to train a traditional machine learning model. Fault data is challenging to obtain because machines operate predominantly in healthy states and are typically shut down for maintenance immediately after a fault occurs. Machines from the same manufacturer depict different behaviours, which encumbers the use of data from other machines to train the machine learning model. Therefore, training a machine learning classifier to label the machine's healthy and faulty states is not a practical solution. Alternatively, the healthy data of machines is abundantly available after installation to train machine learning models. Normal behaviour modelling is the most suitable solution for the fault detection application in wind turbines. The Normal Behaviour Model (NBM) is a machine learning model trained on data collected during the normal operation of a machine, system, or process. It represents the expected, healthy, or typical behaviour. By learning the normal behaviour, the model detects potential faults by identifying deviations from the expected behaviour. Chesterman et al. (2023) presents a comprehensive overview of normal behaviour modelling techniques for SCADA-based condition monitoring in wind turbines. The NBMs are divided into two categories: statistical NBMs and machine learning NBMs. The statistical NBMs are based on statistical distribution and assume that the expected behaviour data follows a distribution, such as the Gaussian distribution. The statistical NMBs are simple and understandable, but cannot capture the complex patterns or non-linear relationships in data. The literature indicates that statistical Normal Behaviour Models (NBMs), such as Principal Component Analysis (PCA) Campoverde et al. (2022); Pozo et al. (2018) and Autoregressive Integrated Moving Averages (ARIMA) Chesterman et al. (2021), are widely applied for condition monitoring based on Supervisory Control and Data Acquisition (SCADA) data. The machine learning NBMs learn the normal behaviour from data. Machine learning NBMs can learn non-linear relationships in the data, but they are difficult to interpret and require more computational resources. Examples of machine learning NBMs include isolation forest, one-class support vector machine, regression models, and autoencoders. The isolation forest Liu et al. (2008) is a tree-based model that recursively splits data points until each instance is isolated. It operates on the principle that faults and anomalies are sparse and behave differently compared to healthy and normal operational data, allowing them to be isolated with fewer partitions. Faults and anomalies are identified based on the predicted shorter path lengths generated by a

https://doi.org/10.5194/wes-2026-83
Preprint. Discussion started: 6 July 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

forest of randomly constructed isolation trees. Isolation forest performs better when trained on both healthy and faulty data; however, it can also detect anomalies when trained solely on healthy data as an NBM. Isolation forest is applied to monitor the wind turbine pitch system McKinnon et al. (2021) and bearing health Sim et al. (2023), using SCADA and vibration data, respectively. The one-class support vector machine Schölkopf et al. (2001) model is trained exclusively on normal, healthy data to define a hypersphere that encloses these points. The data points predicted outside this hypersphere boundary are classified as anomalies. It is applied to detect wind turbine gearbox bearing faults from vibration signals Saari et al. (2019) and SCADA data Tutivén et al. (2022). Dao (2022) proposed a multiple linear regression-based NBM for wind turbine gearbox condition monitoring using SCADA data. The model predicts generator speed based on gearbox and generator bearing temperature; the Chow test is employed to provide statistical evidence for rejecting the null hypothesis, indicating potential faults in the wind turbine. Jamil et al. (2025) proposes a regression-based machine learning model to label individual signal-processing indicators in each operating regime as healthy, warning, or faulty using SCADA and vibration data. The labelled indicators are then fused to provide a high-level health overview of wind turbine drivetrain components. An autoencoder is a deep-learning NBM which learns to compress input data into a lower-dimensional representation and then reconstruct the original input from this compressed representation. The autoencoder is an unsupervised method, and if trained only on healthy data, it develops a bias to reconstruct healthy data. Consequently, a normal-behaviour autoencoder identifies faults and anomalies by evaluating the reconstruction error: the model reconstructs faulty input data as healthy due to the learned bias to reconstruct healthy output, resulting in higher reconstruction errors that signal potential faults. Perez-Sanjines et al. (2023) and Helsen et al. (2023) proposed a normal-behaviour autoencoder method trained from a spectral coherence map derived from vibration signals to detect wind turbine gearbox faults. Autoencoders trained on raw vibration signals Miele et al. (2022) and SCADA data Chen et al. (2021); Renström et al. (2020); Zhao et al. (2018) are used for wind turbine condition monitoring. Previous research on physics-informed multivariate autoencoders employed time-domain statistical Jamil et al. (2023a) and frequency-domain Jamil et al. (2023b) signal processing indicators for detecting wind turbine gearbox faults.

2 Multivariate Autoencoder

The proposed method is a physics-informed normal behaviour model designed as an automated pipeline to simplify the condition monitoring process. This approach leverages a multivariate autoencoder capable of processing large volumes of data, specifically, condition indicators derived from vibration signals, to provide a high-level assessment of the health status of mechanical systems. The model is trained using healthy data collected during the normal operation of mechanical systems, allowing it to learn the system's healthy behaviour. Once trained, the model continuously monitors incoming data and identifies faults by generating alerts whenever it detects deviations from the learned normal behaviour. This high-level condition indicator replaces the need to individually monitor multiple condition-monitoring indicators and simplifies the decision-making process for operators. By offering a concise and automated health overview, the method enables operators to plan efficient maintenance strategies, reducing downtime and improving system productivity. The workflow of the proposed method is divided into key

https://doi.org/10.5194/wes-2026-83
Preprint. Discussion started: 6 July 2026

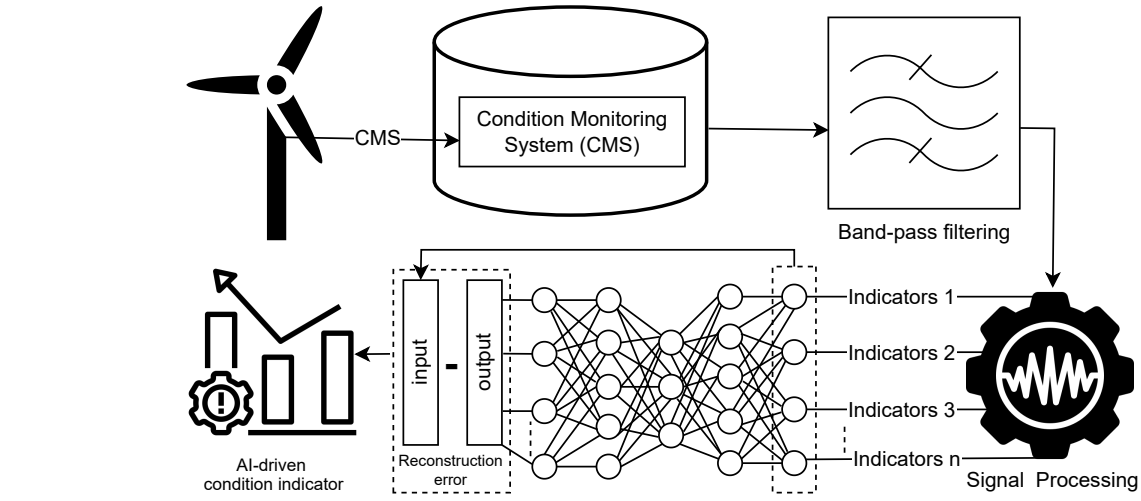


Figure 1. The hybrid condition monitoring pipeline commences by measuring vibration data from wind turbine components. The data is then processed to compute physics-based signal processing indicators, which are used as input to the multivariate autoencoder. The autoencoder's output is used to predict alarms related to potential faults.

2.1 Data measurement and signal processing

To enable the subsequent monitoring pipeline, an in-house developed data acquisition system is employed that allows for measuring long-duration, high-sample-rate data, such that the state-of-the-art algorithms can be fully exploited for dealing with the challenges of slow-rotating machinery like wind turbines.

The signal processing stage of the pipeline consists of multiple steps to deal with the complexity and variability in experimental vibration data Peeters et al. (2016). First off, any new measurement data is checked for its data quality, which can be application dependent, e.g. certain energy levels should be met, and specific spectral signatures are expected to be present in the data. Bypassing this data quality check risks unnecessary computational cost. The second step is typically to estimate the instantaneous angular speed or rotation speed of one of the machine components Peeters et al. (2017); Protopapadakis et al. (2025); Peeters et al. (2022). This speed is not only necessary to properly calculate frequency-based indicators but is also often used to resample the time-domain vibration signal into the angular domain to compensate for any speed variations, be it minor or major ones. After the angular resampling, the signal is separated into stochastic and deterministic signal components. This is done since these bearing and gear signals produce mainly vibration signals that pertain to one of those two categories, with the former usually producing stochastic signals and the latter deterministic signals. Given that deterministic signals produce

https://doi.org/10.5194/wes-2026-83
Preprint. Discussion started: 6 July 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

discrete frequency peaks in the spectrum, this can be used to enable automated signal separation using techniques such as discrete-random separation Peeters et al. (2025), self-adaptive noise cancellation Antoni and Randall (2004), cepstral editing Peeters et al. (2018), linear prediction filtering Peeters et al. (2019a); Kestel et al. (2023); Peeters et al. (2024). After this signal separation, which essentially doubles the number of signals you measure per sensor, the next step is to filter your signal into many different frequency bands to enhance the frequency-based sensitivity of the condition indicators computed on the different filtered signals. Depending on the use case, different filtering approaches can be chosen. In this work, a filtering approach similar to the kurtogram Antoni (2007) is chosen. This approach filters the signal according to a binary-ternary filterbank and results in 24 narrow-band signals that span different frequency bands from 0 Hz to the Nyquist frequency. Finally, a set of statistics is computed on the resulting pre-processed signals. While it is possible to compute hundreds of different statistics on a time series signal, most of them are highly correlated and quantify in some way either the energy level, the non-Gaussianity, or the non-stationarity of the signal Antoni et al. (2024a). Therefore, this work limits the number of computed statistical indicators to seven per pre-processed signal. These eight statistics include: RMS, kurtosis, crest factor, peak energy index, time negentropy, and spectral negentropy. Computing all these indicators for each pre-processed signal, thus quickly results in a very large indicator set per sensor and per machine to monitor. Hence, monitoring a large fleet of machines, each with several sensors installed, quickly leads to several thousand trends to be monitored Helsen et al. (2017); Jamil et al. (2022); Peeters et al. (2024). Not only is this task highly impractical for any human expert, but it also does not allow for taking into account any non-linearities that might be present in the indicator trends. This practical issue thus necessitates the use of more automated methodologies that can deal with such a large indicator set. The next section goes into more concrete detail as to how machine learning can deal with this issue.

2.2 Deep learning

Deep learning models consist of numerous interconnected computational units, referred to as neurons, organised across multiple processing layers. This hierarchical architecture enables the models to learn intricate data representations over multiple levels of abstraction LeCun et al. (2015). The proposed method is a normal behaviour multivariate autoencoder designed to aggregate multiple input condition indicators into a single high-level health indicator output.

2.2.1 Autoencoder

An autoencoder is a neural network employed for different unsupervised learning tasks such as data compression, dimensionality reduction, feature engineering, anomaly detection, image and audio denoising, image reconstruction, and generative tasks Berahmand et al. (2024). The autoencoder is designed to reconstruct input data after compressing it into a latent space representation Rumelhart et al. (1986). This is achieved through a two-step process: encoding the input into a compressed latent representation and subsequently decoding it back to its original form, as illustrated in Figure 2. It consists of three parts: an encoder, a latent space, and a decoder. The encoder transforms the input data into a compressed representation, referred to as the latent space. The latent space is a lower-dimensional, compressed representation of the input data that captures its most important features. The decoder takes the compressed representation from the latent space and reconstructs an approximation

https://doi.org/10.5194/wes-2026-83
Preprint. Discussion started: 6 July 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

of the original input data. The autoencoder architecture learns by gradually reducing the dimensionality of its hidden layers, emphasising the most significant features to uncover the underlying structure and patterns within the input data. This systematic compression is essential to establish a meaningful latent space representation to ensure the model avoids a direct one-to-one mapping between input and output features.

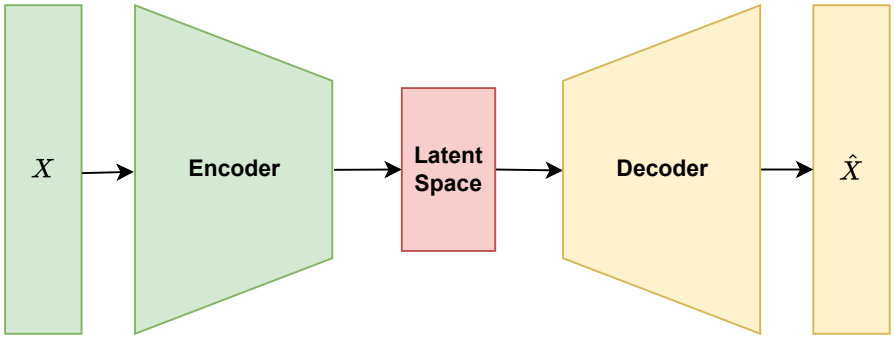


Figure 2. The schematic diagram of the deep autoencoder comprises three key components: the encoder, the latent space, and the decoder. The input data, X, is passed through the input layer to the encoder, which compresses the input to the latent space, and the decoder reconstructs the output, $\hat{X}$, to closely approximate the input by minimising the reconstruction error.

The proposed normal behaviour multivariate autoencoder method is trained using healthy data collected during the machine's expected normal operations. The healthy period is identified by experts based on the mean time between failures, a reliability specification criterion Engelhardt and Bain (1986), or through monitoring condition indicators trends. The selection of training data is critical to ensuring the reliability of the normal behaviour model, as a significant presence of outliers or faulty data can distort the model's understanding of normal behaviour, potentially causing it to overlook faults. However, the model is robust enough to tolerate a small percentage of outliers or faulty data during training. After the careful selection of healthy data D, which is a time series of M observations, denoted as t_i with $i = 1, ..., M$.

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

Each timestamp t_i within the time series is a set of Ncomputed indicators over an individual vibration signal measurement.

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

The architecture of the proposed method is shown in Figure 3, which provides an overview of the deep autoencoder and its components. The exact autoencoder architecture varies depending on the number of computed condition indicators used as input to the model. A feedforward autoencoder using a fully connected linear-layer architecture is employed; however, combining time-domain statistical indicators with frequency-domain indicators significantly increases the input size, leading to a substantial number of trainable parameters in the network. A one-dimensional convolutional neural network (1D-CNN) architecture is utilised to reduce computational complexity, significantly lowering the number of parameters required compared to a fully connected feedforward neural network. The network architecture consists of three main components: an encoder, a

https://doi.org/10.5194/wes-2026-83
Preprint. Discussion started: 6 July 2026

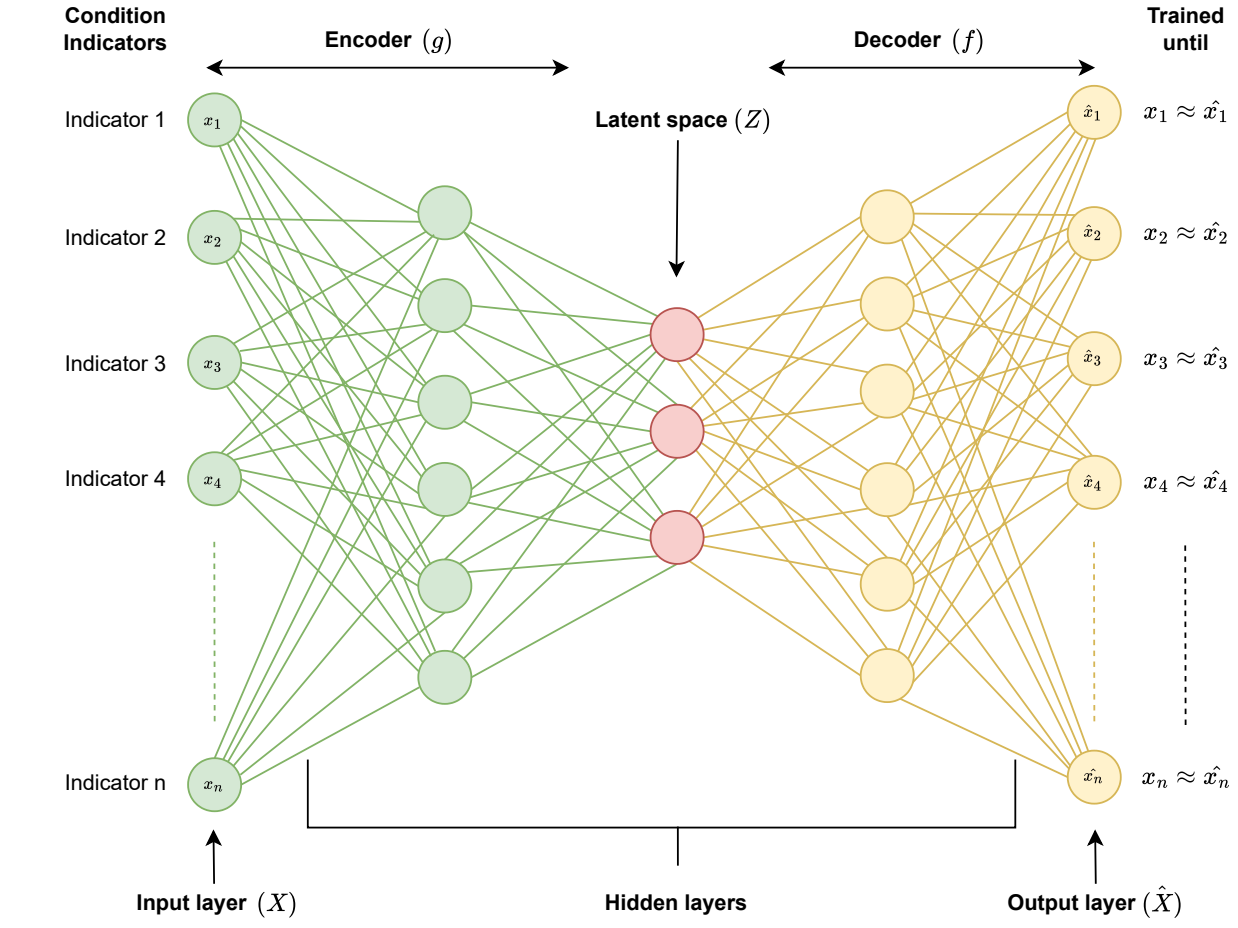


Figure 3. The schematic diagram of the multivariate autoencoder architecture. The condition indicators are input to the encoder, which compresses the input into a latent space representation. The decoder then reconstructs the output from this latent space by closely approximating the input while minimising the reconstruction error.

latent space, and a decoder. The encoder, represented by the function g, maps the input X_i to a lower-dimensional latent representation Z_i. This transformation is mathematically described by Equation 3:

$$Z_i = g(X_i) \tag{3}$$

The decoder represented by function f reconstructs the original input $\hat{X}_i$ from the latent representation Z_i using the following equation **??**:

$$\hat{X}_i = f(Z_i) \tag{4}$$

https://doi.org/10.5194/wes-2026-83
Preprint. Discussion started: 6 July 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

Where the $\hat{X}_i$ is reconstructed representation of the original input X_i:

$$\hat{X}_i = \{\hat{x}_1, \hat{x}_2, \hat{x}_3, ...\hat{x}_N\} \tag{5}$$

The autoencoder is trained using backpropagation, an iterative process to adjust the weights of the neural network parameters to minimise reconstruction error, the difference between the original input and its reconstructed output. It is achieved by computing the error through a loss function and propagating it backwards through the network layers while updating the weights using an optimisation algorithm. The mean squared error (MSE) serves as the evaluation metric for the reconstruction error between the original input X_i and the reconstructed output $\hat{X}_i$:

$$RE = MSE = \frac{1}{N}\sum_{i=1}^{N}(X_i - \hat{X}_i)^2 \tag{6}$$

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. A regularisation term is added to the loss function to avoid overfitting.

$$L_2(W) = \lambda\sum_{i=1}^{n} W_i^2 \tag{7}$$

W represents the set of neural network parameters, n is the total number of neural network parameters, and λ is the regularisation parameter to control the effect of regularisation. $L2$ regularisation, also referred to as weight decay, penalises the magnitude of the weights by adding the sum of the squared neural network parameters to the loss function. This penalty term discourages the optimisation algorithm from assigning large values to the weights, which reduces the model's complexity and improves generalisation. Adam optimisation algorithm Kingma (2014) is used during the training process to iteratively adjust the model parameters, aiming to identify optimal values to minimise the loss function. The training continues until the model achieves minimal reconstruction error, indicating the ability to accurately reconstruct the healthy condition indicator trends with minimal error:

$$X : \{x_1, x_2, x_3, ...x_n\} \approx \hat{X} : \{\hat{x}_1, \hat{x}_2, \hat{x}_3, ...\hat{x}_n\} \tag{8}$$

The trained model, exclusively using healthy data, inherently exhibits a bias towards reconstructing healthy behaviour. As a result, when presented with faulty measurements, it struggles to reconstruct them accurately. In such cases, the reconstruction error, defined as the difference between the faulty input and its reconstructed output, serves as a high-level indicator of the machine's health status. A larger reconstruction error values indicate a higher likelihood of anomalous conditions and potential faults.

The high-level health indicator is represented as a point in a multidimensional space defined by the condition indicators. Figure 4 illustrates a simplified three-dimensional representation with three condition indicators, x_1, x_2, and x_3, where the high-level condition indicator corresponds to a weighted combination of these indicators. While Figure 4 demonstrates the concept with only three indicators for clarity, in practice, the model incorporates hundreds or even thousands of condition indicators. Additional indicators enable the model to compute a more comprehensive high-level health indicator by leveraging the

https://doi.org/10.5194/wes-2026-83
Preprint. Discussion started: 6 July 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

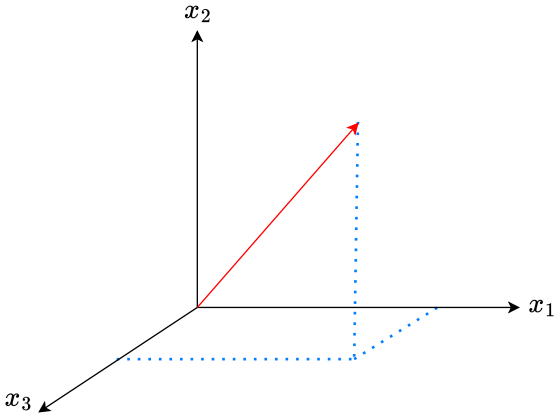


Figure 4. simplified high-level representation of a condition indicator in three dimensions, illustrating an example derived from three signal processing indicators: x_1, x_2, and x_3.

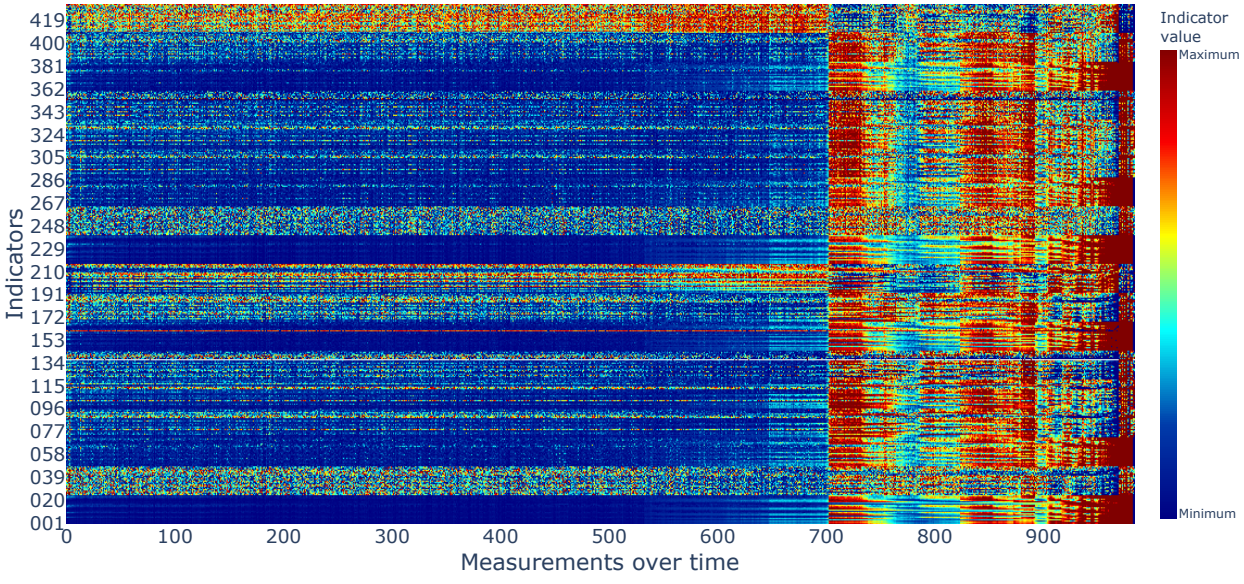


Figure 5. The heatmap of time-domain statistical indicators over time. The x-axis represents the measurements over time, while the y-axis corresponds to the signal processing indicators.

https://doi.org/10.5194/wes-2026-83
Preprint. Discussion started: 6 July 2026

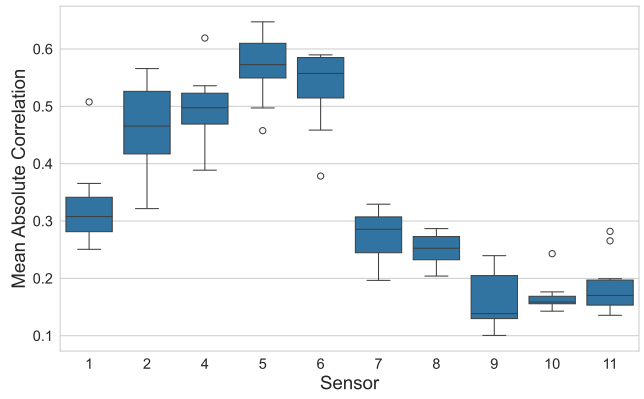
eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

collective information derived from the complete set of indicators. Figure 5 presents a comprehensive set of scaled time-domain statistical indicators computed on a publicly available rolling element-bearing dataset from NASA's Intelligent Maintenance Systems (IMS) Teubert (2004) using a heatmap. A heatmap visually represents data through varying colour intensities to emphasise patterns or anomalies within the dataset. In the figure, the condition indicators for each vibration signal measurement are displayed over time along horizontal lines. Blue regions indicate stable trends with minimum indicator values, while red regions highlight fault trends in the indicators.

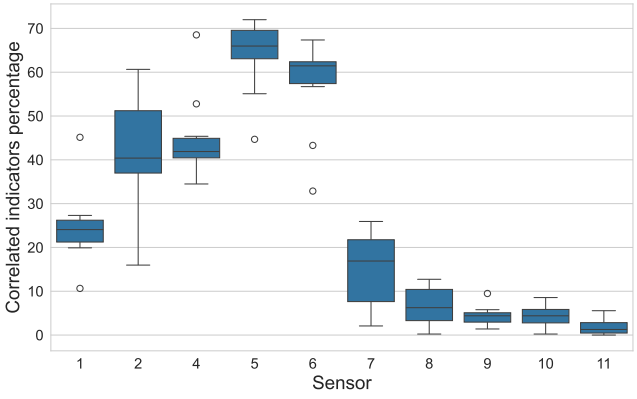
2.2.2 SCADA free condition monitoring

The proposed multivariate autoencoder is a deep, nonlinear model which is proficient in learning the intricate relationships among various condition indicators. By training on these numerous indicators, the model can inherently incorporate the effects of changing operating conditions, since some indicators are strongly correlated with them. It provides a significant advantage, as it allows the model to be trained exclusively on vibration signal-based indicators, eliminating the need for retrieving operating conditions from SCADA data. It is a key difference from linear normal behaviour models, which rely on operating conditions from SCADA data to predict the health status of mechanical components Jamil et al. (2025). A correlation analysis between condition indicators and operating conditions is performed to establish the basis for the proposed method, motivating the use of a non-linear deep learning model to capture these interdependencies and detect faults without explicitly relying on operating conditions data. The study employs a dataset from 10 offshore wind turbines, each equipped with 11 accelerometers to capture the vibration signatures of various drivetrain components. To investigate the relationship between these vibration-based condition indicators and operating conditions, two methods were used. First, a direct correlation computation is performed between the condition indicators of a single sensor and the corresponding operating conditions. Second, a supervised deep learning model is trained to predict operating conditions from the condition indicators computed from a single sensor's vibrational signal. The model's training data are sourced from the same healthy operational period, mirroring the dataset used for the training of the multivariate autoencoder. Two operating conditions, active power and rotational speed, are used to examine the relationship between operating conditions and vibration-based condition indicators. These same two operating conditions are also employed in the linear normal behaviour fault detection method Jamil et al. (2025). Figure 6a presents the boxplot of mean absolute correlation, calculated across 10 wind turbines, between the operating conditions and the condition indicators for each sensor. The results show that particular sensors' vibration indicators, based on the mounted location, exhibit stronger correlations. The consistent mean absolute correlation across the 10 wind turbines indicates that the influence of operating conditions varies depending on the sensor location. Similarly, Figure 6b shows the percentage of vibration indicators that are highly correlated with operating conditions, defined as having a mean absolute correlation greater than 0.5. The boxplot indicates that sensor vibration signals with strong influence from operating conditions exhibit a higher percentage of highly correlated indicators. A deep learning regression model is trained on the vibrational condition indicators to demonstrate the capability of a non-linear model in predicting operating conditions, such as generator speed and active power. Figure 7 presents the model's goodness of fit (R^2) on the test data for generator speed and active power across 11 sensors, where each value represents the median performance over 10 wind turbines. The R^2 metric, ranging from 0 to 1,

https://doi.org/10.5194/wes-2026-83
Preprint. Discussion started: 6 July 2026

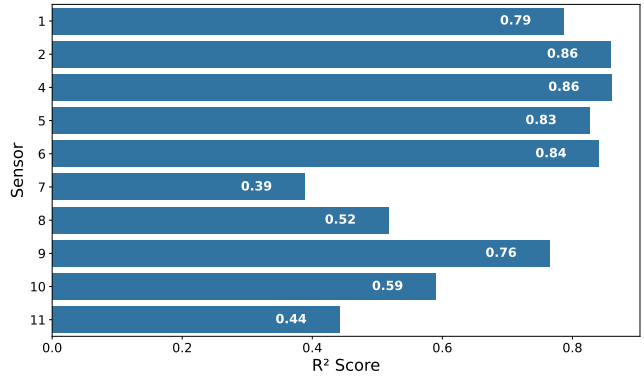


(a) Mean absolute correlation.

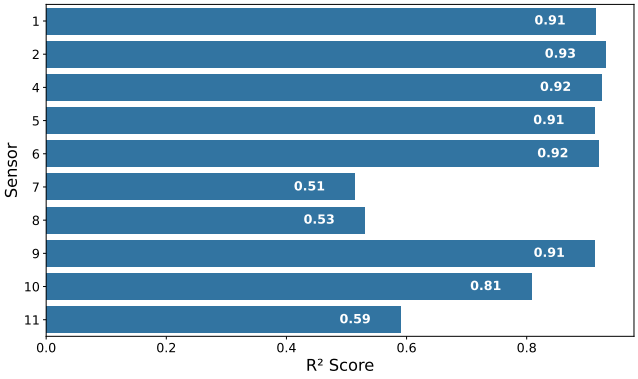


(b) vibrational indicator percentage with correlation > 0.5.

Figure 6. Boxplots summarising the correlation between vibration indicators and operating conditions across 10 wind turbines, each equipped with 11 sensors. Subfigure (a) reports the mean absolute correlation values for each sensor, while subfigure (b) reports the percentage of indicators with a mean absolute correlation greater than 0.5, computed across all turbines.



(a) Generator speed goodness of fit



(b) Active power goodness of fit

Figure 7. Regression model goodness of fit (R^2) for predicting generator speed and active power using vibration-based condition indicators across 11 sensors. The results show median values over 10 wind turbines.

quantifies how well the model's predictions match the actual data. The results show that regression models built on vibration indicators from highly correlated sensors can more accurately predict operating conditions. Together, the correlation analysis and the regression model performance confirm that vibration signatures exhibit strong correlation with operating conditions, with the degree of influence depending on the sensor's location. These findings demonstrate that the proposed multivariate autoencoder inherently incorporates the effect of changing operating conditions from the vibration-based condition indicators. Consequently, SCADA operating conditions data is not required for training the normal-behaviour multivariate autoencoder

https://doi.org/10.5194/wes-2026-83
Preprint. Discussion started: 6 July 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

for fault detection, nor is it necessary for predictions, which can be very useful for remote edge computing Verstraeten et al. (2019).

2.3 Threshold computation and fault alarm generation

Fault alarms are generated using z-scores, a statistical measure that quantifies the deviation from the learned normal behaviour. The z-score z_i at each input measurement is computed as shown in Equation 9.

$$z_i = \frac{RE_i - \mu}{\sigma} \tag{9}$$

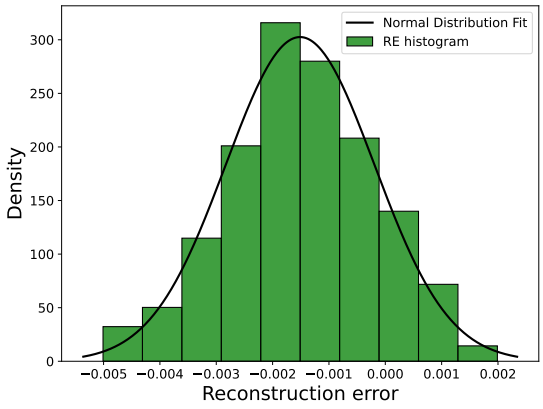
Where RE_i is the reconstruction error of the respective input, and μ and σ represent the mean and standard deviation, respectively. These statistical values are computed from the predicted reconstruction errors by the trained normal behaviour autoencoder model on the healthy training dataset D, as shown in Equation 10 and Equation 11.

$$\mu = \frac{1}{M}\sum_{i=1}^{M} RE_i \tag{10}$$

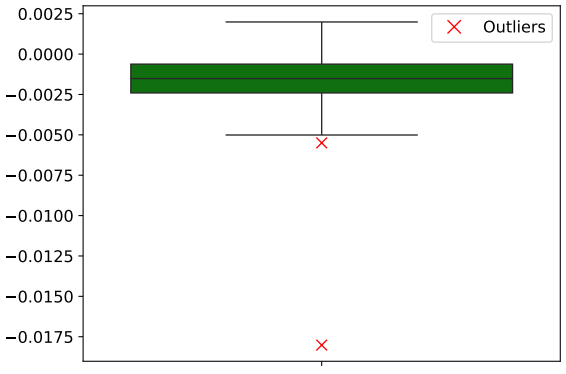
$$\sigma = \sqrt{\frac{1}{M}\sum_{i=1}^{M}(RE_i - \mu)^2} \tag{11}$$

Before computing the mean and standard deviation, a normality test is performed on the predicted reconstruction error, and potential outliers are removed to ensure the data fits a normal distribution. Figure 8a illustrates the normal distribution of the IMS healthy dataset for the initial 400 measurement predictions. The two outliers identified in Figure 8b are removed using the interquartile range method Smiti (2020) to achieve a better fit for the data. Two threshold levels, corresponding to z-score values of $T_1 = 3$ and $T_2 = 5$, are used to classify faults into warning and alarm states, respectively. The z-score indicates how many standard deviations the measured value deviates from the mean normal behaviour. The selection of the threshold can vary across different applications; therefore, the k-sigma rule can be adopted, where the threshold is defined as K standard deviations (σ) from the mean (μ) of the normal behaviour. The trained model evaluates the health status of mechanical components by continuously analysing live measured data in real-time. The z-score, computed from the reconstruction error (the difference between the actual and the predicted values), provides a high-level health indicator of the mechanical component's health status. During normal mechanical operations, the model predicts the data with a minimum error, resulting in low reconstruction errors and correspondingly low z-scores. However, in the presence of a fault or anomaly, prediction errors increase, leading to high z-score values. Each newly acquired data point is categorised based on z-score thresholds: If the z-score is less than 3, the data point is classified as healthy. If the z-score lies between 3 and 5, it is marked as a warning. If the z-score exceeds 5, the data point is labelled as an alarm. To reduce false alarms caused by measurement noise or transient anomalies during healthy periods, a sliding window approach is applied. The window moves across the z-score timeseries, replacing each z-score at a given timestamp with the median value within the current window. The sliding window only incorporates the current and past values, ensuring that the method is suitable for real-time deployment.

https://doi.org/10.5194/wes-2026-83
Preprint. Discussion started: 6 July 2026



(a) The normal distribution fit on the autoencoder predictions of healthy data.



(b) Outlier removal from autoencoder predictions on healthy data to fit the normal distribution.

Figure 8. Outlier removal from autoencoder predictions on healthy data before computing mean and standard deviation.

3 Experiments

The proposed method is validated using NASA's intelligent maintenance systems (IMS) bearing fault dataset provided by the University of Cincinnati Teubert (2004). This dataset is widely recognised for benchmarking the research in bearing fault detection, condition monitoring, and predictive maintenance, as it contains healthy and faulty data collected under controlled and well-documented experimental conditions Sacerdoti et al. (2023); Gousseau et al. (2016). The experimental setup consists

Dataset	Measurements	Channels	Experiment Duration (min)	Recorded Signal Duration (min)	Announced Damages
1	2,156	8	49,680	36	Bearing 3: Inner race Bearing 4: Rolling element
2	984	4	9,840	16	Bearing 1: Outer race
3	4,448	4	44,480	74	Bearing 3: Outer race

Table 1. IMS dataset description.

of a test rig with four Rexnord ZA-2115 double-row bearings installed on a rotating shaft. The shaft's rotational speed is maintained at a constant 2000 RPM using an AC motor coupled to the shaft via rubber belts. A radial load of 6000 lbs is applied to the shaft and bearings through a spring mechanism, and all bearings are force-lubricated. PCB 353B33 high-sensitivity quartz

https://doi.org/10.5194/wes-2026-83
Preprint. Discussion started: 6 July 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

ICP accelerometers are mounted on the bearing housings to capture the vibration signals. The data is sampled at a rate of 20

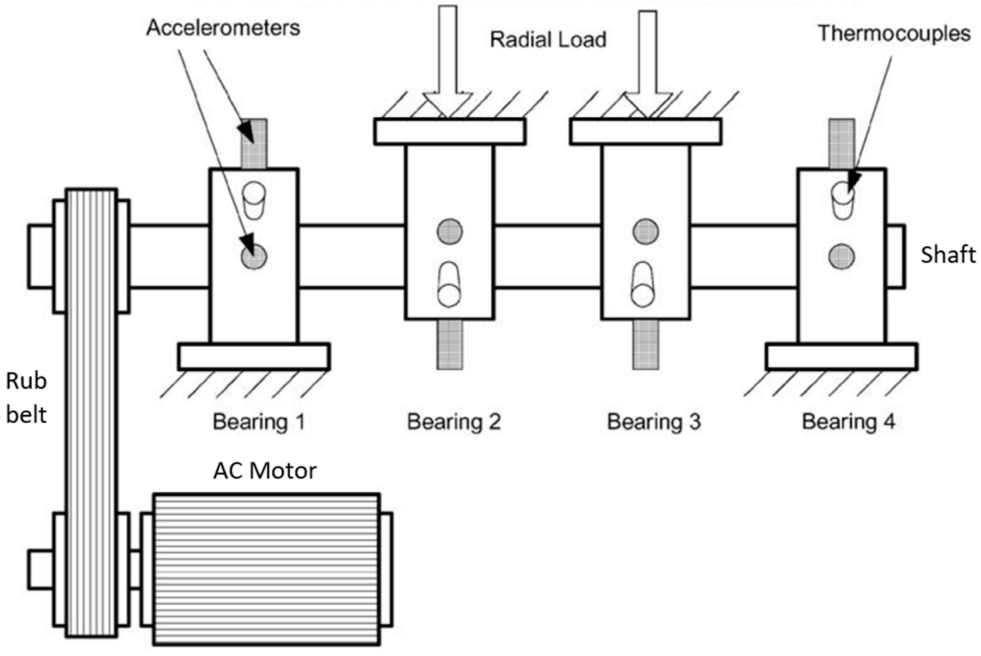


Figure 9. IMS experimental setup for bearing fault dataset.

kHz for one second every ten minutes. The description of the dataset is provided in Table 1, for dataset 1, two accelerometers per bearing are used to measure vibrations along the x and y axes, while for datasets 2 and 3, only one accelerometer per bearing is mounted, as illustrated in Figure 9. The dataset includes three additional datasets, each measured during independent test-to-failure experiments. Dataset 1 features inner race faults in Bearing 3 and roller element faults in Bearing 4. Dataset 2 contains an outer race fault in Bearing 1, while Dataset 3 features an outer race fault in Bearing 3. This dataset provides a robust foundation for evaluating the effectiveness of the proposed method. In addition to benchmarking on the publicly available IMS dataset, the proposed method is also validated using real-world data from an offshore wind farm. Due to confidentiality agreements, specific details about the wind turbines cannot be disclosed. The wind farm consists of numerous multi-megawatt turbines, with vibration signals recorded from various components using accelerometers mounted on the gearbox housings. Ten-second vibration signals were captured at intervals of two to three days over several years, providing a comprehensive dataset for analysis. Additionally, multiple fault cases are labelled based on the existing monitoring methods Jamil et al. (2025), with each fault subsequently verified by on-site technicians, ensuring reliable ground truth for evaluation.

https://doi.org/10.5194/wes-2026-83
Preprint. Discussion started: 6 July 2026

4 Results

The proposed multivariate autoencoder method is validated using two distinct datasets. The open-source public IMS dataset is employed to benchmark and validate the method, while the offshore wind farm dataset demonstrates the real-life application. The results are presented in Figure 10, Figure 11, Figure 12, and Figure 13. The green overlay indicates the healthy training period, while the solid red vertical line marks the measurement at which the fault is detected. The dashed yellow and red horizontal lines represent the warning and alarm thresholds, respectively. Measurements are classified as green, yellow, or red based on these thresholds, corresponding to healthy, warning, and alarm states. During the healthy state of the component, the predicted high-level health indicator remains below the warning threshold, and the measurements are labelled green, indicating the normal behaviour. In the presence of a fault, the high-level health indicator deviates from normal behaviour. If it exceeds the warning but remains below the alarm threshold, the measurements are labelled yellow, indicating a warning state. As fault severity increases, the high-level health indicator surpasses the alarm threshold, and measurements are labelled red, indicating an alarm state and a fault in the respective component. False alarms, caused by measurement errors or noise in the data, are labelled healthy despite exceeding threshold levels. These false alarms are mitigated by applying a sliding window, as described in the methodology to filter out the outliers.

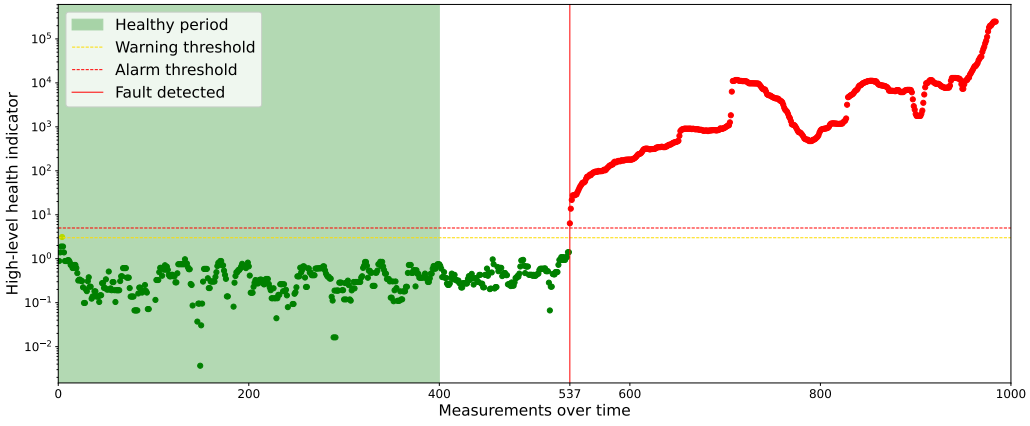


Figure 10. The IMS dataset bearing outer race fault trend prediction. The x-axis represents the measurements over time, and the y-axis corresponds to the log scale of the high-level health indicator. The green overlay represents the healthy data used to train the model. The colours green, yellow, and red indicate the healthy, warning, and alarm states, respectively.

4.1 IMS bearing fault dataset results

The second IMS dataset is used to validate the proposed method. It contains an outer race fault in Bearing 1 and comprises 984 measurements of 1 second at a 20 kHz sampling rate, recorded at 10-minute intervals. Since the fault is located in bearing 1,

https://doi.org/10.5194/wes-2026-83
Preprint. Discussion started: 6 July 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

channel 1 measurements are used to compute signal processing indicators. As shown in the histogram of indicators in Figure 5, the initial 400 measurements represent the healthy period and are used to train the multivariate autoencoder. The trained model is then applied to predict the health of bearing 1 in all 984 measurements. The fault is detected at the 537th measurement and remains consistently identified until the bearing fails, as illustrated in Figure 10. The green overlay indicates the healthy training period, while the solid red vertical line marks the measurement at which the fault is detected. The dashed yellow and red horizontal lines represent the warning and alarm thresholds, respectively. Measurements are classified as green, yellow, or red on the basis of the thresholds, corresponding to healthy, warning, and alarm states.

4.2 Wind farm drivetrain dataset results

The proposed method is evaluated on data from an offshore wind farm with more than 50 wind turbines. Due to confidentiality

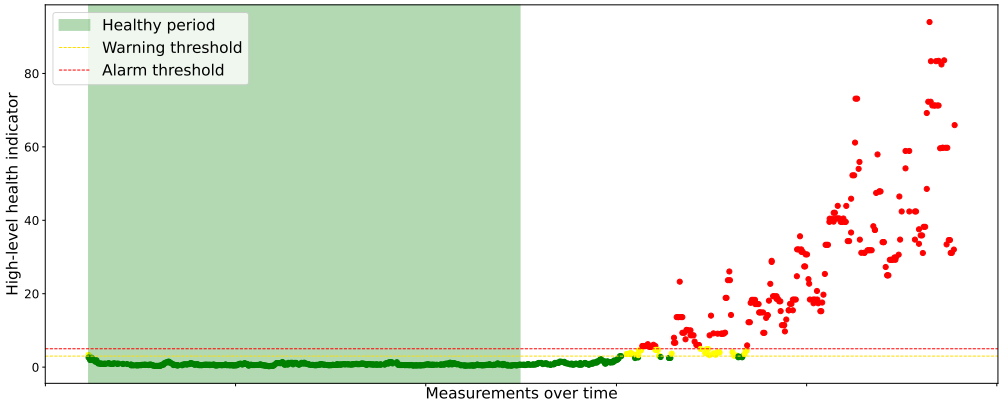


Figure 11. The wind turbine generator channel fault trend prediction. The x-axis represents the measurements over time, and the y-axis corresponds to the high-level health indicator. The green overlay represents the healthy data used to train the model. The colours green, yellow, and red indicate the healthy, warning, and alarm states, respectively.

agreements, specific details of the wind turbines are not disclosed, and the results are presented with anonymised axes. Multiple accelerometers are installed on the wind turbine drivetrain to capture the vibration signatures of the components. A single multivariate autoencoder is trained on the computed condition indicators derived from the vibration signature of a specific channel. The trained model is then used to monitor the health of the component on which the accelerometer is mounted. The model is trained using two years of healthy data and is subsequently used to monitor the health of wind turbine drivetrain components, including the generator, high-speed stage, intermediate stage, planetary stage, and main bearing, over several years. Examples of fault detection cases from the offshore wind farm dataset are presented in Figure 11, Figure 12, and Figure 13. Furthermore, the faults predicted by the proposed method are confirmed through manual borescope inspections.

https://doi.org/10.5194/wes-2026-83
Preprint. Discussion started: 6 July 2026

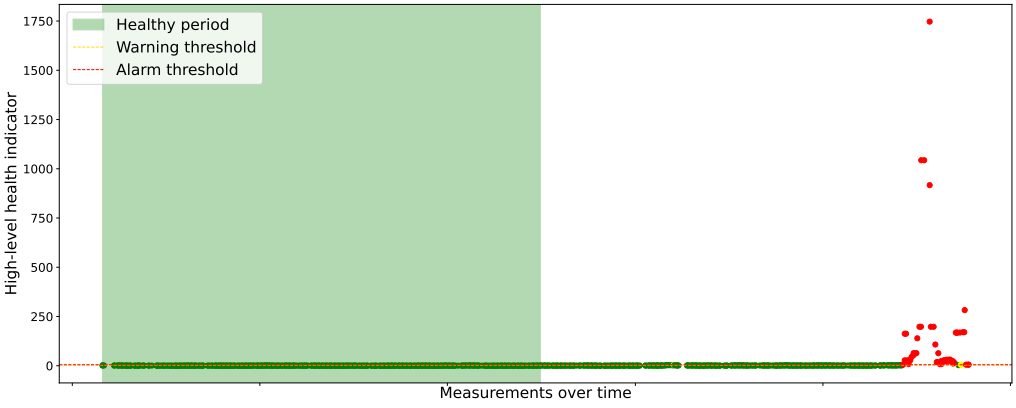


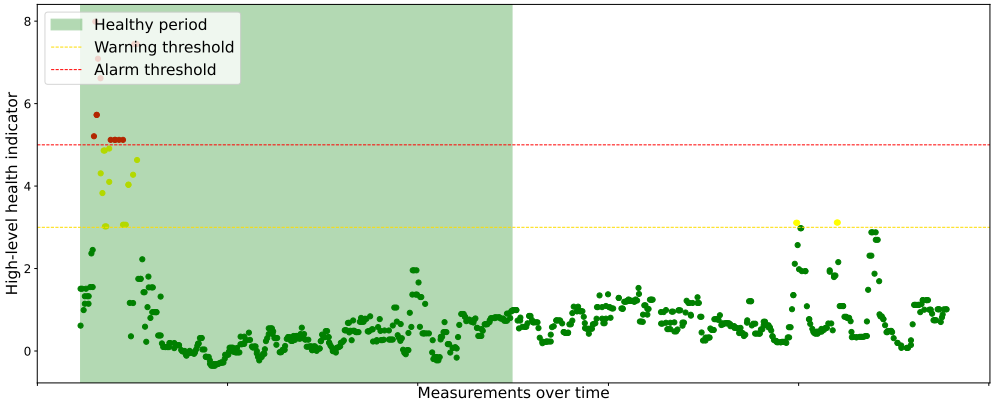
Figure 12. The wind turbine planetary stage gear fault trend prediction. The x-axis represents the measurements over time, and the y-axis corresponds to the high-level health indicator. The green overlay represents the healthy data used to train the model. The colours green, yellow, and red indicate the healthy, warning, and alarm states, respectively.

The fault in the wind turbine generator is depicted in Figure 11. The model predictions remain in a healthy state after the training period but begin to deviate from the normal behaviour after the introduction of a fault, and the model prediction gradually increases with the severity of the fault. The planetary stage gear fault is critical due to its high potential to develop into severe damage. Early detection of such faults is essential. However, in most cases, the condition indicators exhibit a steep rise in fault trends, making early detection more challenging. A planetary stage gear fault is illustrated in Figure 12; the model is trained on condition indicators computed from the planetary stage channel's vibration signature. The trained model can predict a steep rise in the alarm trend as a result of the fault. The proposed method combines numerous condition indicators to produce a single high-level artificial intelligence-based indicator, addressing the varying sensitivities of different condition indicators to specific fault types. A planetary stage gear fault case, shown in Figure 13, illustrates a scenario where the planetary stage gear's fault is evident in the frequency domain but not in the time domain. As depicted in Figure 13a, a model relying only on time-domain statistical indicators fails to predict the fault. However, by integrating both time-domain and frequency-domain condition indicators, the model successfully detects the fault as shown in Figure 13b. Moreover, certain condition indicators can detect specific faults at an early stage, while others identify them only after the faults have progressed. However, the proposed method is capable of detecting such faults early by leveraging the condition indicators that exhibit early fault trends. This further validates the model's capability to identify faults even when only a subset of indicators captures the fault trend.

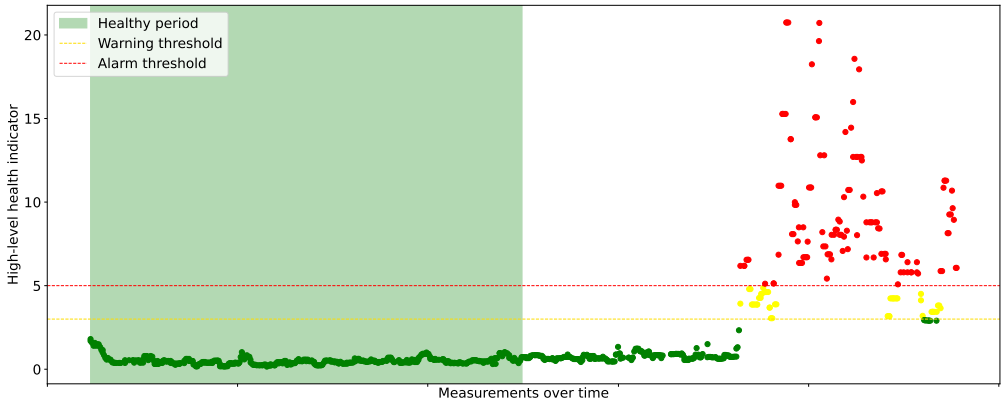
4.3 Performance analysis

The performance of the proposed fault detection method is validated using data from 10 wind turbines, each equipped with 11 vibration sensors mounted on various drivetrain components. Manual inspections confirmed faults in 19 drivetrain components

https://doi.org/10.5194/wes-2026-83
Preprint. Discussion started: 6 July 2026



(a) Time domain statistical indicators.



(b) Time and frequency domain indicators.

Figure 13. The wind turbine planetary stage gear fault trend prediction, comparing fault detection using time-domain statistical indicators with the combination of both time-domain and frequency-domain indicators. The x-axis represents the measurements over time, and the y-axis corresponds to the high-level health indicator. The green overlay represents the healthy data used to train the model. The colours green, yellow, and red indicate the healthy, warning, and alarm states, respectively.

across 8 of these turbines. Vibration signatures from a single faulty component can propagate through the drivetrain and be detected by sensors on adjacent components; therefore, the false alarm analysis is restricted to the two remaining healthy wind turbines. The confusion matrix is used to evaluate the performance of the model, using the following definitions for true positives (TPs), true negatives (TNs), false positives (FPs), and false negatives (FNs):

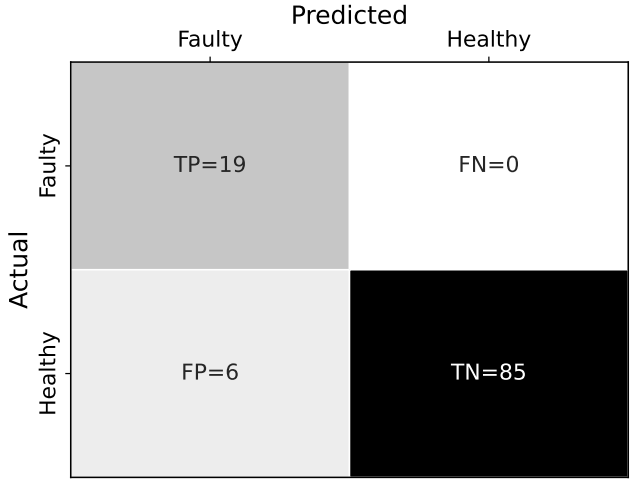
- TP: the model correctly identifies a manually confirmed fault as faulty.

https://doi.org/10.5194/wes-2026-83
Preprint. Discussion started: 6 July 2026

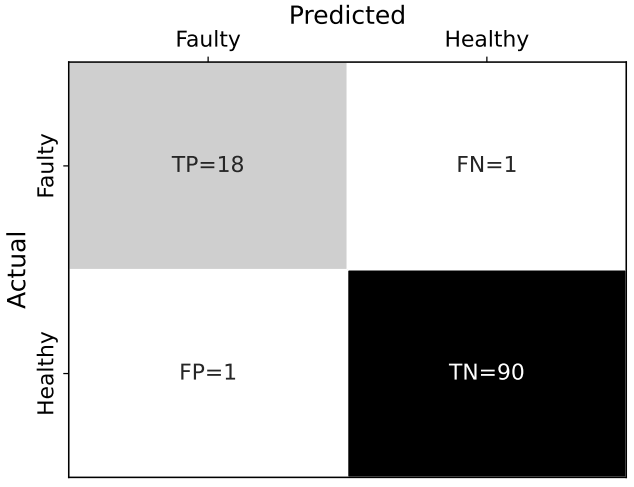
eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

- TN: the model correctly classifies a healthy case as healthy.
- FP (false alarm): the model incorrectly classifies a healthy case as faulty.
- FN (missed fault detection): the model fails to detect a fault, incorrectly classifying it as healthy.

To evaluate the model's sensitivity and specificity, two sets of threshold levels are applied. The first (lower) threshold uses z-score values of 3 for warnings and 5 for alarms. In contrast, the second (higher) threshold increases these values to 5 and 10, respectively, providing a more conservative fault classification. The confusion matrix derived from the lower threshold level,



(a) Confusion Matrix with low threshold.



(b) Confusion Matrix with high threshold.

Figure 14. Confusion matrices for low (LT) and high (HT) threshold levels, derived from 10 offshore wind turbines, each equipped with 11 vibration sensors.

as shown in Figure 14a, shows 19 TPs, 0 FN (missed fault detection), 6 FPs (false alarms), and 85 TNs. The lower threshold prioritises early fault detection by enhancing the model's sensitivity to confirmed fault cases. While the model successfully identifies all 19 faults confirmed through manual inspection, it also incorrectly classifies 6 healthy cases as faulty, leading to reduced specificity. The higher threshold level improves the model's specificity. As depicted in Figure 14b, the resulting confusion matrix reports 18 TPs, 1 FN (missed fault detection), 1 FPs (false alarms), and 90 TNs. The validation dataset has 19 manually confirmed fault cases, while the remaining 91 cases are considered healthy. Due to this class imbalance, standard performance metrics such as precision, recall, and F1-score cannot provide an accurate assessment of the model's performance. Therefore, weighted precision Equation 12, recall Equation 13, and F1-score Equation 14 are calculated to provide a comparative assessment of the model's performance.

$$\text{Weighted Precision} = \frac{(N_f \times \text{Precision}_f) + (N_h \times \text{Precision}_h)}{N_f + N_h} \tag{12}$$

https://doi.org/10.5194/wes-2026-83
Preprint. Discussion started: 6 July 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

$$\text{Weighted Recall} = \frac{(N_f \times \text{Recall}_f) + (N_h \times \text{Recall}_h)}{N_f + N_h} \tag{13}$$

$$\text{Weighted F1-Score} = \frac{(N_f \times \text{F1}_f) + (N_h \times \text{F1}_h)}{N_f + N_h} \tag{14}$$

Where N_f and N_h are the number of actual faulty and healthy cases, respectively. Similarly, $Precision_f$, $Recall_f$, and $F1_f$ represent the precision, recall, and F1-score for the faulty label, respectively. On the other hand, $Precision_h$, $Recall_h$, and $F1_h$ represent the precision, recall, and F1-score for the healthy label, respectively.

	Precision LT	**Recall LT**	**F1-score LT**	**Precision HT**	**Recall HT**	**F1-score HT**
Faulty	0.76	1.00	0.86	0.95	0.95	0.95
Healthy	1.00	0.93	0.97	0.99	0.99	0.99
Weighted	**0.96**	**0.95**	**0.95**	**0.98**	**0.98**	**0.98**

Table 2. Weighted precision, recall, and F1-score for low (LT) and high (HT) threshold levels.

The low threshold level achieves a perfect recall of 1.00 for faulty cases, indicating that all confirmed fault instances are successfully detected. However, the high sensitivity results in a lower precision of 0.76 due to false alarm cases. In contrast, the high threshold level improves the overall accuracy to 98%, compared to the 95% achieved with the low threshold. Despite this increase in overall accuracy, the balanced accuracy, as defined in Equation 15, which accounts for the class imbalance, remains consistent at 97% for both levels. This suggests that while the low threshold level is optimal for incipient fault detection, the high threshold level reduces the false positives.

$$\text{Balanced Accuracy} = \frac{1}{2}\left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP}\right) \tag{15}$$

The high threshold provides a more reliable detection strategy by significantly reducing false alarms. However, it also tends to delay fault detection, as alarms are triggered later in the fault progression. This limitation can restrain early-stage fault detection. Consequently, there is a trade-off between early detection achieved through the low threshold and more accurate fault confirmation by the high threshold.

5 Discussion

The research presents a physics-informed deep learning normal behaviour model to facilitate condition monitoring of mechanical components. This approach employs a multivariate autoencoder that fuses multiple signal-processing condition indicators into a single high-level health indicator to provide a comprehensive overview of the health of mechanical components. By incorporating numerous condition indicators into one unified indicator, the method significantly reduces the time and resources

https://doi.org/10.5194/wes-2026-83
Preprint. Discussion started: 6 July 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

required for condition monitoring. The model is trained on healthy vibration data to learn the normal behaviour of mechanical components and detect deviations from normal behaviour when faults occur. A correlation analysis on 10 wind turbines shows that certain vibration-based condition indicators are highly correlated with operating conditions. Furthermore, a multivariate deep learning regression model trained on these indicators explains a high percentage of the variance in operating conditions, depending on sensor sensitivity to operational variations. The proposed non-linear deep learning method leverages the correlation to incorporate the effect of changing operating conditions by capturing the intricate relationships among vibrational condition indicators. As a result, the model can be trained exclusively on vibration data, without relying on operating conditions data, to monitor the health of complex machines under changing operating conditions, such as wind turbines. The proposed method is validated using two datasets: the publicly available IMS dataset, which ensures the reproducibility of results, and an offshore wind farm dataset, which demonstrates the real-life application of the method. The results indicate that different indicators are effective at identifying various types of faults, and the proposed method's ability to fuse numerous condition indicators enables the detection of the majority of fault types. As illustrated in Figure 13, the model successfully detected fault trends when frequency-domain indicators were combined with time-domain statistical indicators as input. This increases the model's input size but improves its capability to detect most faults. The proposed method provides an alternative to traditional condition monitoring approaches, which rely on manually tracking individual indicators and become inadequate for managing large fleets of complex machines. By using condition indicators as input to the model, a single high-level indicator can be derived to streamline the monitoring. The proposed method depends on condition indicators to detect faults. If a fault trend is not visible in any input condition indicator, the model predicts healthy behaviour, as shown in Figure 13a. In this case, the fault is not present in the time domain, and a model trained only on time-domain statistical indicators fails to predict the fault trend. However, incorporating a diverse set of condition indicators mitigates this issue. The proposed method is evaluated using a dataset from 10 wind turbines, each equipped with 11 vibration sensors, comprising a total of 110 sensors and 19 manually confirmed fault cases. The method's sensitivity and specificity are assessed under two threshold levels. At the lower threshold, the model achieves the highest sensitivity at the cost of low specificity by detecting all 19 confirmed fault cases, including 6 false alarms. In contrast, the higher threshold improves specificity by reducing false alarms to 1; however, it misses one confirmed fault and delays detection. Overall, the high threshold achieves an accuracy of 98%, compared to 95% for the low threshold; however, the balanced accuracy is 97% for both levels. These results suggest the threshold level can be adapted to application requirements, while considering the sensitivity–specificity trade-off.

6 Conclusion

This study presented a novel data-driven condition-monitoring method to enhance the efficiency of health monitoring and maintenance for large fleets of complex machines. The proposed method incorporated numerous condition indicators, including time-domain statistical and frequency-domain indicators, to derive a comprehensive high-level health indicator. The method was validated on two datasets: NASA's IMS bearing fault dataset, ensuring the reproducibility of results, and an offshore wind farm dataset, demonstrating the application in real-world scenarios. The findings indicated that by integrating a diverse range of

https://doi.org/10.5194/wes-2026-83
Preprint. Discussion started: 6 July 2026

condition indicators, the method effectively detected the majority of faults. Additionally, it simplified the condition-monitoring process by significantly reducing the time required for analysis, making it particularly suitable for large fleets, such as wind farms. The proposed method, evaluated on a dataset of 10 wind turbines, achieved up to 97% balanced accuracy for fault detection. The results showed that threshold levels could be adapted to application requirements, depending on the preferred balance between sensitivity and specificity. Future work could focus on improving the interpretability of the approach by incorporating explainable AI techniques, further improving its practicality and adoption in industrial settings.

Data availability. The datasets used in this study include both proprietary and public sources. The vibration and operational data from offshore wind farms are proprietary to industrial partners and were obtained under a non-disclosure agreement; therefore, they cannot be shared publicly. To ensure reproducibility and method validation, the publicly available NASA IMS bearing dataset is utilised as a benchmark for fault detection. This public dataset is accessible through the NASA Data Repository Teubert (2004).

Author contributions. FJ was responsible for conceptualisation, data curation, formal analysis, investigation, methodology development, project administration, funding acquisition, software coding, validation, visualisation, and writing of the original draft. CP contributed to conceptualisation, methodology development, project administration, supervision, and the reviewing and editing of the manuscript. JH was responsible for conceptualisation, funding acquisition, methodology development, supervision, and the final phase of paper review and editing.

Competing interests. The contact author has declared that none of the authors has any competing interests

7 Acknowledgements

The authors would like to acknowledge FWO (Fonds Wetenschappelijk Onderzoek) for their support through the SB grants of Faras Jamil (#1S63123N), and SBO CORE project. Furthermore, this research was supported by funding from the Flemish Government under the “Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen” program and under the VLAIO/Blue Cluster Supersized 5.0 ICON project (HBC.2024.0130). The authors are also grateful to the VSC Supercomputing Flanders centre for the support in the context of the VSC Cloud program.

During the preparation of this manuscript, the authors used generative AI and writing-assistant tools, including Grammarly, to proofread and improve the language of the text. These tools were employed solely to streamline the writing process and were not used to generate or interpret scientific content. All AI-assisted revisions were reviewed, edited, and verified by the authors to ensure accuracy and scientific integrity.

https://doi.org/10.5194/wes-2026-83
Preprint. Discussion started: 6 July 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

References

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

Antoni, J. and Randall, R.: Unsupervised noise cancellation for vibration signals: part I—evaluation of adaptive algorithms, Mechanical Systems and Signal Processing, 18, 89–101, https://doi.org/10.1016/S0888-3270(03)00012-8, 2004.

Antoni, J. and Randall, R.: The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines, Mechanical Systems and Signal Processing, 20, 308–331, https://doi.org/10.1016/j.ymssp.2004.09.002, 2006.

Antoni, J., Kestel, K., Peeters, C., Leclère, Q., Girardin, F., Ooijevaar, T., and Helsen, J.: On the design of Optimal Health Indicators for early fault detection and their statistical thresholds, Mechanical Systems and Signal Processing, 218, 111 518, https://doi.org/10.1016/j.ymssp.2024.111518, 2024a.

Antoni, J., Kestel, K., Peeters, C., Leclère, Q., Girardin, F., Ooijevaar, T., and Helsen, J.: On the design of Optimal Health Indicators for early fault detection and their statistical thresholds, Mechanical Systems and Signal Processing, 218, 111 518, https://doi.org/10.1016/j.ymssp.2024.111518, 2024b.

Berahmand, K., Daneshfar, F., Salehi, E. S., Li, Y., and Xu, Y.: Autoencoders and their applications in machine learning: a survey, Artificial Intelligence Review, 57, 28, https://doi.org/10.1007/s10462-023-10662-6, 2024.

Campoverde, L., Tutivén, C., Vidal, Y., and Benaláazar-Parra, C.: Scada data-driven wind turbine main bearing fault prognosis based on principal component analysis, in: Journal of Physics: Conference Series, vol. 2265, p. 032107, IOP Publishing, https://doi.org/10.1088/1742-6596/2265/3/032107, 2022.

Chen, H., Liu, H., Chu, X., Liu, Q., and Xue, D.: Anomaly detection and critical SCADA parameters identification for wind turbines based on LSTM-AE neural network, Renewable Energy, 172, 829–840, https://doi.org/10.1016/j.renene.2021.03.078, 2021.

Chesterman, X., Verstraeten, T., Daems, P.-J., Nowé, A., and Helsen, J.: Condition monitoring of wind turbines and extraction of healthy training data using an ensemble of advanced statistical anomaly detection models, in: Annual Conference of the PHM Society, vol. 13, https://doi.org/10.36001/phmconf.2021.v13i1.2980, 2021.

Chesterman, X., Verstraeten, T., Daems, P.-J., Nowé, A., and Helsen, J.: Overview of normal behavior modeling approaches for SCADA-based wind turbine condition monitoring demonstrated on data from operational wind farms, Wind Energy Science, 8, 893–924, https://doi.org/10.5194/wes-8-893-2023, 2023.

Dao, P. B.: Condition monitoring and fault diagnosis of wind turbines based on structural break detection in SCADA data, Renewable Energy, 185, 641–654, https://doi.org/10.1016/j.renene.2021.12.051, 2022.

Engelhardt, M. and Bain, L. J.: On the Mean Time between Failures for Repairable Systems, IEEE Transactions on Reliability, 35, 419–422, https://doi.org/10.1109/TR.1986.4335491, 1986.

Gousseau, W., Antoni, J., Girardin, F., and Griffaton, J.: Analysis of the Rolling Element Bearing data set of the Center for Intelligent Maintenance Systems of the University of Cincinnati, in: CM2016, Charenton, France, https://hal.science/hal-01715193, 2016.

Helsen, J., Peeters, C., Doro, P., Ververs, E., and Jordaens, P. J.: Wind farm operation and maintenance optimization using big data, in: 2017 IEEE Third International Conference on big data computing service and applications (BigDataService), pp. 179–184, IEEE, https://doi.org/10.1109/BigDataService.2017.27, 2017.

https://doi.org/10.5194/wes-2026-83
Preprint. Discussion started: 6 July 2026

Helsen, J., Perez, F., Jamil, F., Antoni, J., and Peeters, C.: Signal Processing Informed Deep Learning for Failure Detection in a Fleet of Multi-Stage Planetary Gearboxes with Limited Knowledge about Characteristic Frequencies, in: AIAC 2023: 20th Australian International Aerospace Congress, Engineers Australia, Melbourne, https://search.informit.org/doi/10.3316/informit.068781576131458, 2023.

Igba, J., Alemzadeh, K., Durugbo, C., and Eiriksson, E. T.: Analysing RMS and peak values of vibration signals for condition monitoring of wind turbine gearboxes, Renewable Energy, 91, 90–106, https://doi.org/10.1016/j.renene.2016.01.006, 2016.

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

Jamil, F., Jara Avila, F., Vratsinis, K., Peeters, C., and Helsen, J.: 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 of *Turbo Expo: Power for Land, Sea, and Air*, p. V014T37A003, American Society of Mechanical Engineers, https://doi.org/10.1115/GT2023-101689, 2023a.

Jamil, F., Peeters, C., Verstraeten, T., and Helsen, J.: Wind turbine drivetrain fault detection using physics-informed multivariate deep learning, in: Surveillance, Vibrations, Shock and Noise, Institut Supérieur de l'Aéronautique et de l'Espace [ISAE-SUPAERO], Toulouse, France, https://hal.science/hal-04166103, 2023b.

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

Kestel, K., Peeters, C., Antoni, J., Leclère, Q., Girardin, F., and Helsen, J.: Informed sparsity-based blind filtering in the presence of second-order cyclostationary noise, Mechanical Systems and Signal Processing, 199, 110 438, https://doi.org/10.1016/j.ymssp.2023.110438, 2023.

Kingma, D. P.: Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980, https://doi.org/10.48550/arXiv.1412.6980, 2014.

LeCun, Y., Bengio, Y., and Hinton, G.: Deep learning, Nature, 521, 436–444, https://doi.org/10.1038/nature14539, 2015.

Liu, F. T., Ting, K. M., and Zhou, Z.-H.: Isolation Forest, in: 2008 Eighth IEEE International Conference on Data Mining, pp. 413–422, https://doi.org/10.1109/ICDM.2008.17, 2008.

Mathew, J. and Alfredson, R. J.: The Condition Monitoring of Rolling Element Bearings Using Vibration Analysis, Journal of Vibration, Acoustics, Stress, and Reliability in Design, 106, 447–453, https://doi.org/10.1115/1.3269216, 1984.

McKinnon, C., Carroll, J., McDonald, A., Koukoura, S., and Plumley, C.: Investigation of isolation forest for wind turbine pitch system condition monitoring using SCADA data, Energies, 14, 6601, https://doi.org/10.3390/en14206601, 2021.

Miele, E. S., Bonacina, F., and Corsini, A.: Deep anomaly detection in horizontal axis wind turbines using Graph Convolutional Autoencoders for Multivariate Time series, Energy and AI, 8, 100 145, https://doi.org/10.1016/j.egyai.2022.100145, 2022.

Mohd Ghazali, M. H. and Rahiman, W.: Vibration analysis for machine monitoring and diagnosis: a systematic review, Shock and Vibration, 2021, 9469 318, https://doi.org/10.1155/2021/9469318, 2021.

Moors, J.: A quantile alternative for kurtosis, Journal of the Royal Statistical Society: Series D (The Statistician), 37, 25–32, https://doi.org/10.2307/2348376, 1988.

Peeters, C., Guillaume, P., and Helsen, J.: Vibration data pre-processing techniques for rolling element bearing fault detection, in: Proceedings of the 23rd international conference on sound & vibration, ISSN 2329-3675, 2016.

Peeters, C., Leclère, Q., Antoni, J., Guillaume, P., and Helsen, J.: Vibration-based angular speed estimation for multi-stage wind turbine gearboxes, Journal of Physics: Conference Series, 842, 012 053, http://stacks.iop.org/1742-6596/842/i=1/a=012053, 2017.

https://doi.org/10.5194/wes-2026-83
Preprint. Discussion started: 6 July 2026

Peeters, C., Guillaume, P., and Helsen, J.: Vibration-based bearing fault detection for operations and maintenance cost reduction in wind energy, Renewable Energy, 116, 74–87, https://doi.org/10.1016/j.renene.2017.01.056, 2018.

Peeters, C., Antoni, J., and Helsen, J.: Blind vibration filtering using envelope linear prediction for fault detection without knowledge of machine kinematics, in: SURVISHNO, vol. 1, pp. 1–10, INSA Lyon, 2019a.

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

Peeters, C., Verstraeten, T., Nowé, A., and Helsen, J.: 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, p. V010T09A014, American Society of Mechanical Engineers, https://doi.org/10.1115/OMAE2019-96713, 2019c.

Peeters, C., Antoni, J., Leclère, Q., Verstraeten, T., and Helsen, J.: Multi-harmonic phase demodulation method for instantaneous angular speed estimation using harmonic weighting, Mechanical Systems and Signal Processing, 167, 108 533, https://doi.org/10.1016/j.ymssp.2021.108533, 2022.

Peeters, C., Wang, W., Blunt, D., Verstraeten, T., and Helsen, J.: Fatigue crack detection in planetary gears: Insights from the HUMS2023 data challenge, Mechanical Systems and Signal Processing, 212, 111 292, https://doi.org/10.1016/j.ymssp.2024.111292, 2024.

Peeters, C., Antoni, J., and Helsen, J.: A multi-delay extension of the Discrete/Random Separation method, Mechanical Systems and Signal Processing, 236, 112 959, https://doi.org/10.1016/j.ymssp.2025.112959, 2025.

Perez-Sanjines, F., Peeters, C., Verstraeten, T., Antoni, J., Nowé, A., and Helsen, J.: Fleet-based early fault detection of wind turbine gearboxes using physics-informed deep learning based on cyclic spectral coherence, Mechanical Systems and Signal Processing, 185, 109 760, https://doi.org/10.1016/j.ymssp.2022.109760, 2023.

Pozo, F., Vidal, Y., and Salgado, : Wind Turbine Condition Monitoring Strategy through Multiway PCA and Multivariate Inference, Energies, 11, https://doi.org/10.3390/en11040749, 2018.

Protopapadakis, G., Peeters, C., Leclère, Q., Antoni, J., and Helsen, J.: Enhancing instantaneous angular speed estimation with an adaptive Multi-Order Probabilistic Approach, Mechanical Systems and Signal Processing, 226, 112 322, https://doi.org/10.1016/j.ymssp.2025.112322, 2025.

Randall, R. B. and Antoni, J.: Rolling element bearing diagnostics—A tutorial, Mechanical Systems and Signal Processing, 25, 485–520, https://doi.org/10.1016/j.ymssp.2010.07.017, 2011.

Renström, N., Bangalore, P., and Highcock, E.: System-wide anomaly detection in wind turbines using deep autoencoders, Renewable Energy, 157, 647–659, https://doi.org/10.1016/j.renene.2020.04.148, 2020.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J.: Learning internal representations by error propagation, Biometrika, 71, 6, https://doi.org/10.7551/mitpress/4943.003.0128, 1986.

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

Sacerdoti, D., Strozzi, M., and Secchi, C.: A Comparison of Signal Analysis Techniques for the Diagnostics of the IMS Rolling Element Bearing Dataset, Applied Sciences, 13, https://doi.org/10.3390/app13105977, 2023.

Schölkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J., and Williamson, R. C.: Estimating the Support of a High-Dimensional Distribution, Neural Computation, 13, 1443–1471, https://doi.org/10.1162/089976601750264965, 2001.

https://doi.org/10.5194/wes-2026-83
Preprint. Discussion started: 6 July 2026

Sharma, V. and Parey, A.: A Review of Gear Fault Diagnosis Using Various Condition Indicators, Procedia Engineering, 144, 253–263, https://doi.org/10.1016/j.proeng.2016.05.131, international Conference on Vibration Problems 2015, 2016.

Sim, J., Min, J., Kim, S., Lee, S. W., and Choi, J.-H.: Construction of bearing health indicator under time-varying operating conditions based on Isolation Forest, Engineering Applications of Artificial Intelligence, 126, 107 058, https://doi.org/10.1016/j.engappai.2023.107058, 2023.

Smiti, A.: A critical overview of outlier detection methods, Computer Science Review, 38, 100 306, https://doi.org/https://doi.org/10.1016/j.cosrev.2020.100306, 2020.

Teubert, C.: IMS Bearings | NASA Open Data Portal — data.nasa.gov, https://data.nasa.gov/dataset/ims-bearings, [Accessed 06-12-2024], 2004.

Tutivén, C., Vidal, Y., Insuasty, A., Campoverde-Vilela, L., and Achicanoy, W.: Early Fault Diagnosis Strategy for WT Main Bearings Based on SCADA Data and One-Class SVM, Energies, 15, https://doi.org/10.3390/en15124381, 2022.

Večeř, P., Kreidl, M., and Šmíd, R.: Condition indicators for gearbox condition monitoring systems, Acta Polytechnica, 45, https://doi.org/10.14311/782, 2005.

Verstraeten, T., Gomez Marulanda, F., Peeters, C., Daems, P.-J., Nowé, A., and Helsen, J.: Edge computing for advanced vibration signal processing, in: Surveillance, Vishno and AVE conferences, INSA-Lyon, Université de Lyon, Lyon, France, https://hal.science/hal-02188766, 2019.

Zhao, H., Liu, H., Hu, W., and Yan, X.: Anomaly detection and fault analysis of wind turbine components based on deep learning network, Renewable Energy, 127, 825–834, https://doi.org/10.1016/j.renene.2018.05.024, 2018.

Zhu, J., Nostrand, T., Spiegel, C., and Morton, B.: Survey of condition indicators for condition monitoring systems, in: Annual Conference of the PHM Society, vol. 6, https://doi.org/10.36001/phmconf.2014.v6i1.2514, 2014.