https://doi.org/10.5194/wes-2025-123
Preprint. Discussion started: 6 August 2025
© Author(s) 2025. CC BY 4.0 License.

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

# Fault Detection in Wind Turbines Using Health Index Monitoring with Variational Autoencoders

Shun Wang[1], Yolanda Vidal[1,2], and Francesc Pozo[1,2]

[1]Control, Data and Artificial Intelligence (CoDAlab), Department of Mathematics, Escola d'Enginyeria de Barcelona Est (EEBE), Universitat Politècnica de Catalunya (UPC), Campus Diagonal-Besòs (CDB), Eduard Maristany, 16, Barcelona, 08019, Spain
[2]Institute of Mathematics (IMTech), Universitat Politècnica de Catalunya (UPC), Barcelona, 08028, Spain

**Correspondence:** Yolanda Vidal (yolanda.vidal@upc.edu)

**Abstract.** As wind energy capacity expands globally, ensuring the operational reliability and economic viability of wind turbines has become a critical industrial challenge. Effective fault detection systems are essential for minimizing high maintenance costs and preventing catastrophic failures. To address this need, this paper presents a semi-supervised framework designed to identify anomalies in wind turbines using only healthy operational data. The methodology begins by extracting a comprehensive set of features from the time and frequency domains of raw vibration signals to capture a rich representation of the dynamics of wind turbines. A variational autoencoder, a deep generative model, is then trained exclusively on these features from healthy operational periods to learn a robust model of normal behavior and generate reconstruction errors as health indicators with exponentially weighted moving average smoothing to enhance robustness and reduce false alarms. The framework is evaluated using public data from the Aventa AV-7 ETH Zurich Research Wind Turbine, which includes multiple failure events. Experimental results demonstrate effective and early detection of pitch faults, as well as accurate detection of icing events and aerodynamic imbalances. The proposed approach therefore offers a robust and practical solution to improving operational safety and enabling proactive maintenance of wind turbines.

## 1 Introduction

Wind energy has become a major renewable technology, with global installed capacity reaching 1,136 GW in 2024 (Council, 2025). Despite this growth, the increasing deployment of wind turbines, particularly in remote and challenging offshore environments, has underscored significant operational and maintenance (O&M) challenges. These challenges directly threaten the economic viability of wind energy projects, as O&M costs can constitute 20-30% of the levelized cost of electricity (Ren et al., 2021; Irena, 2018).

Wind turbines operate in highly dynamic and stressful conditions, experiencing aerodynamic, gravitational, centrifugal, and gyroscopic loads (Badihi et al., 2022). These harsh operational environments, combined with the remote locations of wind installations, make turbines particularly susceptible to component failures and premature degradation. Consequently, there is an urgent and critical need for effective condition monitoring and predictive maintenance strategies to improve turbine reliability and reduce operational expenditures.

https://doi.org/10.5194/wes-2025-123
Preprint. Discussion started: 6 August 2025
© Author(s) 2025. CC BY 4.0 License.

Prognostics and health management (PHM) has become an indispensable framework to optimize wind turbine maintenance (Cuesta et al., 2025). PHM encompasses three core tasks: fault detection (identifying when a fault occurs), fault diagnosis (determining the type and location of the fault) and prognostics (predicting the remaining useful life of the components) (Cuesta et al., 2025). Among these, fault detection represents the foundation for effective maintenance strategies, enabling early identification of component degradation before failures occur.

Modern wind turbines are equipped with both supervisory control and data acquisition (SCADA) systems and high-frequency condition monitoring systems (CMS). While SCADA systems provide low-frequency (e.g., 10-minute intervals) operational data, CMS, particularly through vibration monitoring, captures high-frequency signals essential for health assessment. Vibration signals contain rich and detailed information about the health of rotating machinery, revealing characteristic fault signatures in components such as blades, gearboxes, and bearings that are undetectable in SCADA data (Ashkarkalaei et al., 2025; Castellani et al., 2024).

Current fault detection approaches are broadly categorized into model-based methods (Habibi et al., 2019), and data-driven approaches (Abid et al., 2021; Chesterman et al., 2023). Model-based methods use relevant professional knowledge to establish physical representations based on first principles, but are often difficult to generalize and require constant updates to reflect real-world operational complexities (Dey et al., 2015; Ashkarkalaei et al., 2025). In contrast, data-driven methods have gained significant traction by learning directly from sensor data without the need for an explicit physical model (Pandit and Wang, 2024; Xu et al., 2024; Rezamand et al., 2020). Early data-driven techniques focused on the use of signal processing to extract statistical features and health indices (HIs) from time-domain (Wang et al., 2025b) and frequency-domain data (Jiang et al., 2011; Ying et al., 2025). More recently, advances in machine learning, and deep learning in particular, have enabled the automatic extraction of complex patterns. Architectures such as convolutional neural networks (Jiang et al., 2018), recurrent neural networks (Yu et al., 2020; Encalada-Dávila et al., 2022), and autoencoders (AEs) (Chen et al., 2021; Wang et al., 2025a) are now widely used to improve fault detection performance.

However, existing data-driven approaches for vibration analysis have predominantly focused on single-channel signals, thereby overlooking the rich, system-level information available in a multi-channel wind turbine setup. Effectively utilizing such data presents several key challenges. First, the high dimensionality and complexity of raw vibration signals can lead to an increased computational burden and model overfitting. Second, the scarcity of labeled fault data in operational settings makes supervised learning impractical. Furthermore, the inherent variability of operating conditions (for example, wind speed) can cause anomaly detection models like traditional AEs to produce false alarms, affecting their reliability. Therefore, a robust fault detection framework must effectively handle high-dimensional data, operate without fault examples, and provide stable indications of degradation.

To address the unique challenges of analyzing multi-channel vibration data, this study introduces an integrated fault detection framework. To manage the high dimensionality and complexity inherent in the raw signals, the framework first employs a feature engineering module. This step extracts a comprehensive set of features from the multi-channel vibration data and concatenates them into a unified, system-level feature vector, making the subsequent learning task more tractable and effective. To overcome the common issue of scarce fault data in industrial settings, our framework then utilizes an semi-supervised

https://doi.org/10.5194/wes-2025-123
Preprint. Discussion started: 6 August 2025
© Author(s) 2025. CC BY 4.0 License.

learning paradigm centered on a variational autoencoder (VAE). By training on data from healthy operations, the VAE learns a robust probabilistic model of the system's normal behavior. Finally, to ensure stable and reliable detection while minimizing false alarms caused by operational variability, the raw reconstruction error from the VAE is processed using an exponentially weighted moving average (EWMA). This post-processing step smooths the resulting health index, effectively filtering transient noise while highlighting persistent trends indicative of true degradation.

The effectiveness of the framework is demonstrated on the Aventa AV-7 ETH Zurich Research Wind Turbine (Chatzi et al., 2023). The experimental results show an early detection of pitch faults and an accurate identification of icing events and aerodynamic imbalances. This work has been developed to participate in the ASCE-EMI Structural Health Monitoring for the Wind Energy Challenge (WeDoWind), part of the WeDoWind RTDT Research Affiliate Program space, and in collaboration with the ASCE Structural Health Monitoring & Control Committee.

## 2 Related Work

This section reviews data-driven fault detection methods for wind turbines, first by explaining the different machine learning paradigms and then by focusing on autoencoder-based models commonly used for anomaly detection.

### 2.1 Learning Paradigms for Wind Turbine Fault Detection

Machine learning and deep learning-based fault detection methods are typically categorized into supervised, unsupervised, and semi-supervised approaches. Supervised methods are powerful but depend on large and accurately labeled datasets that represent both *normal* and various *faulty* states (Dibaj et al., 2023; Yang et al., 2024). However, these methods face significant practical limitations in the wind energy domain due to the diverse and unpredictable nature of faults and the immense difficulty of obtaining a comprehensive, labeled dataset for every possible failure mode (Rezamand et al., 2020). Collecting such data on-site is both challenging and time-consuming, limiting the applicability of supervised approaches.

Unsupervised learning operates on entirely unlabeled datasets, attempting to identify anomalies without any prior knowledge of the turbine's health status. The core assumption is that faults are rare and structurally distinct from the majority of the data. Models such as the isolation forest (Xu et al., 2023) or clustering-based algorithms work by identifying these *lonely* points that are few and different. However, this poses a challenge in complex operational settings. The primary limitation is the inability to differentiate between a genuine, subtle fault and a benign but infrequent operational state (e.g., a specific startup sequence or emergency stop). Since the model lacks a defined concept of *normal*, any rare event can be flagged as an anomaly, which can lead to a high false alarm rate.

Semi-supervised learning, particularly methods focused on normal behavior modeling, presents a pragmatic and effective compromise. In this paradigm, a model is trained exclusively on healthy turbine operational data, which allows it to develop a detailed understanding of normal behavior (Chesterman et al., 2023). These approaches avoid the need for labeled fault data, relying on the assumption that anomalies will manifest as deviations from normal operational patterns. This paradigm includes various strategies including regression-based (Encalada-Dávila et al., 2022; Bilendo et al., 2023), reconstruction-based (Jin

https://doi.org/10.5194/wes-2025-123
Preprint. Discussion started: 6 August 2025
© Author(s) 2025. CC BY 4.0 License.

et al., 2024; Wu et al., 2023; Chen et al., 2023), and one-class classification methods (Tutivén et al., 2022). Among these methods, autoencoder-based reconstruction methods, which learn to recreate healthy data from a compressed representation, have become particularly prominent.

### 2.2 Autoencoder-Based Anomaly Detection and Health Index Construction

Within the semi-supervised framework, autoencoders and their variants have become prominent models for anomaly detection. The fundamental principle is to train a model to accurately reconstruct its input data. When trained exclusively on healthy data, the AE learns a compressed representation of normal operational patterns. Consequently, it will exhibit a low reconstruction error (RE) for healthy, seen data but a significantly higher RE for unseen, anomalous data. This RE is a powerful and widely used foundation for building a health index, where increasing RE values signal potential degradation (Jiang et al., 2017).

The research community has progressively improved AE architectures to improve the performance of wind turbine fault detection. For example, Jiang et al. (Jiang et al., 2017) employed a denoising autoencoder, while Chen et al. (Chen et al., 2020) used stacked denoising autoencoders to improve robustness. Wang et al. (Wang et al., 2022) proposed an improved autoencoder specifically designed to improve the accuracy of fault detection. More complex models have also been developed by integrating AEs with other deep learning techniques. For example, long short-term memory-autoencoder (LSTM-AE) networks (Chen et al., 2021) capture temporal dependencies in the data, enhancing the model's ability to detect subtle fault patterns over time. Similarly, convolutional neural network-conditional variational autoencoder models (CNN-CVAE) (Liu et al., 2022) and long short-term memory-stacked denoising autoencoders (LSTM-SDAE) (Zhang et al., 2022) have been proposed to further improve detection capabilities.

Despite these architectural advances, significant challenges persist, particularly when applying these models to the high-dimensional, multi-channel vibration data addressed in this paper. First, the direct application of these models to high-dimensional raw vibration signals can obscure fault signatures and increase the complexity of the model. Second, the raw HI derived from the reconstruction error is often volatile and susceptible to fluctuations from varying operational conditions, leading to false alarms. To address these, the proposed method begins with targeted feature engineering to distill robust, low-dimensional inputs from raw signals. Then a variational autoencoder is employed, which is a probabilistic generative model known for its ability to learn a smooth, well-regularized representation of normal behavior. Finally, a post-processing step is used, applying an exponentially weighted moving average to the VAE's reconstruction errors for a stable and reliable HI.

## 3 Methodology

This section provides the key components of the framework, including feature extraction, variational autoencoder, and the application of exponentially weighted moving average for anomaly detection.

https://doi.org/10.5194/wes-2025-123
Preprint. Discussion started: 6 August 2025
© Author(s) 2025. CC BY 4.0 License.

## 3.1 Overview

The general structure of the framework is illustrated in Figure 1. It begins by extracting features from raw vibration signals, which capture critical time- and frequency-domain characteristics. The extracted features are then inputted to the VAE, which encodes the input into a latent space that is regularized to approximate a multivariate normal distribution $\mathcal{N}(0, I)$. The reconstruction process within the VAE is guided by a loss function that combines the reconstruction loss and Kullback-Leibler (KL) divergence, ensuring accurate reconstruction and effective latent space regularization.

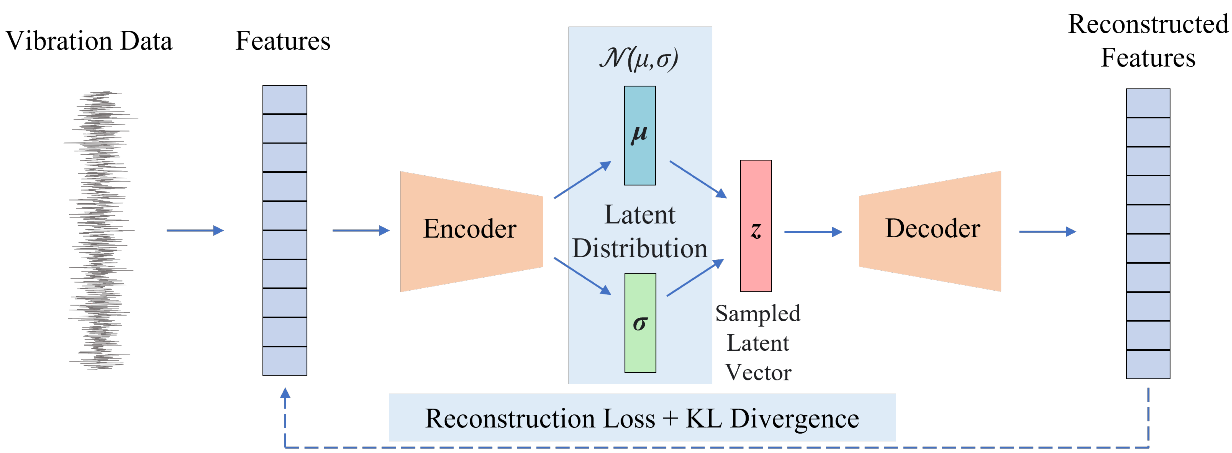


**Figure 1.** An overview of the proposed framework. Initially, the time domain and frequency domain features of a single vibration signal is first extracted. The extracted features are then encoded into a latent distribution characterized by a mean vector ($\mu$) and a standard deviation vector ($\sigma$). A latent vector $z$ is sampled from this distribution and passed through the decoder to reconstruct these features. The reconstruction process is guided by a loss function that combines reconstruction loss and Kullback-Leibler (KL) divergence.

## 3.2 Feature Extraction

Firstly, feature extraction is performed on each vibration sample. A total of 20 distinct time- and frequency-domain features are extracted to characterize the vibration signals accurately, as illustrated in Table 1. These features are selected based on their demonstrated effectiveness in structural health monitoring and fault detection, as evidenced by previous studies (Ding et al., 2021; Li et al., 2024; Sandoval et al., 2021; Zhou et al., 2022), demonstrating their effectiveness in detecting problems in rotating machinery and wind turbine components. Specifically, time-domain features provide insight into statistical and physical properties, such as signal variability, energy, and peaks, which are essential for identifying transient faults and irregular patterns. On the other hand, frequency-domain features capture the spectral characteristics of the vibration signal, enabling the identification of fault-specific harmonics and energy distributions. By integrating features from both domains, a holistic representation of the signal is achieved, enhancing the robustness and reliability of anomaly detection.

https://doi.org/10.5194/wes-2025-123
Preprint. Discussion started: 6 August 2025
© Author(s) 2025. CC BY 4.0 License.

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

**Table 1.** Formulae for the selected features in time and frequency domains.

| **Domain** | **Features and Formulae** | |
|---|---|---|
| **Time domain** | Standard deviation $\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(a_i - \mu)^2}$ | Root mean square $\text{RMS} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} a_i^2}$ |
| | Maximum value $A_{\max} = \max(\lvert a \rvert)$ | Minimum value $A_{\min} = \min(a)$ |
| | Peak-to-peak value $A_{pp} = \max(a) - \min(a)$ | Skewness $S = \frac{1}{n}\sum_{i=1}^{n}\frac{(a_i-\mu)^3}{\sigma^3}$ |
| | Kurtosis $K = \frac{1}{n}\sum_{i=1}^{n}\frac{(a_i-\mu)^4}{\sigma^4} - 3$ | Waveform factor $\text{WF} = \frac{\text{RMS}}{\text{mean}(\lvert a \rvert)}$ |
| | Peak factor $\text{PF} = \frac{A_{\max}}{\text{RMS}}$ | Impulse factor $\text{IF} = \frac{A_{\max}}{\text{mean}(\lvert a \rvert)}$ |
| | Absolute Mean value $\text{AM} = \frac{1}{n}\sum_{i=1}^{n}\lvert a_i \rvert$ | Crest factor $C_f = \frac{A_{\max}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}}$ |
| | Clearance factor $C_l = \frac{A_{\max}}{(\frac{1}{n}\sum_{i=1}^{n}\sqrt{\lvert x_i \rvert})^2}$ | |
| **Frequency domain** | Spectral mean $F_\mu = \frac{1}{m}\sum_{i=1}^{m}\lvert F_i \rvert$ | Spectral variance $F_{\sigma^2} = \frac{1}{m}\sum_{i=1}^{m}(\lvert F_i \rvert - F_\mu)^2$ |
| | Spectral std. deviation $F_\sigma = \sqrt{F_{\sigma^2}}$ | Spectral entropy $H_F = -\sum_{i=1}^{m} p_i \log_2(p_i)$ |
| | Spectral energy $E_F = \frac{\sum_{i=1}^{m}\lvert F_i \rvert^2}{m}$ | Spectral skewness $S_F = \frac{1}{m}\sum_{i=1}^{m}\left(\frac{\lvert F_i \rvert - F_\mu}{F_\sigma}\right)^3$ |
| | Spectral kurtosis $K_F = \frac{1}{m}\sum_{i=1}^{m}\frac{(\lvert F_i \rvert - F_\mu)^4}{{F_\sigma}^4} - 3$ | |

*Note:* $\{a_i\}$ represents vibration signals, $n$ is the the total amount of sampling points, $\mu$ represents the mean value of vibration signal, and $\sigma$ is the standard deviation. $\{F_i\}$ represents frequency components, $m$ is the number of frequency components, and $p_i = \frac{\lvert F_i \rvert}{\sum_{k=1}^{m}\lvert F_k \rvert}$.

### 3.3 Variational Autoencoder

Following the extraction of features, the resulting features, now denoted as $x$, are processed using a variational autoencoder to model the feature distribution of normal operational data. The VAE framework consists of two main components: an encoder, $Q_\phi(x)$, which maps features $x$ into a latent distribution $z$, and a decoder, $P_\theta(z)$, which reconstructs the original features $x$ from the latent codes $z$. The structure of the VAE model is illustrated in Figure 1.

Unlike traditional deterministic autoencoders that learn point-to-point mappings, VAEs employ a probabilistic framework that models data distributions rather than individual data points (Doersch, 2016; Kingma et al., 2013). This probabilistic nature makes VAEs particularly suitable for fault detection applications where there are inherent signal variations even in healthy machinery due to factors such as load fluctuations, speed variations, or measurement noise (Yan et al., 2021). Traditional autoencoders minimize reconstruction loss through deterministic mapping. In contrast, VAEs learn the underlying probability distribution $p_{\text{data}}(x)$ using latent variables $z$ through variational inference. The encoder $Q_\phi(x)$ outputs parameters of a mul-

https://doi.org/10.5194/wes-2025-123
Preprint. Discussion started: 6 August 2025
© Author(s) 2025. CC BY 4.0 License.

Open Access WIND ENERGY SCIENCE DISCUSSIONS

tivariate Gaussian distribution (mean $\mu(x)$ and variance $\sigma^2(x)$), while the decoder $P_\theta(z)$ reconstructs features from sampled latent codes. This probabilistic approach enables better handling of data variability.

The VAE optimizes the evidence lower bound (ELBO) through two complementary loss terms: reconstruction loss and KL
150 divergence loss. The reconstruction loss measures the difference between the original input data and its reconstruction from the latent space. This loss ensures that the VAE accurately captures the underlying structure of the features. The reconstruction loss is defined as:

$$\mathcal{L}_{\text{rec}} = \frac{1}{n}\sum_{i=1}^{n} \|x_i - \hat{x}_i\|_2^2, \tag{1}$$

where $x_i$ represents the $i$-th input sample, $\hat{x}_i$ is the reconstructed sample, and $n$ is the total number of samples. This loss
155 encourages the encoder $Q$ and decoder $P$ to work together to reconstruct the input $x$ as accurately as possible.

In addition to the reconstruction loss, the VAE incorporates a KL divergence loss, which ensures that the learned latent codes $z$ approximate a prior distribution, typically a multivariate normal distribution $\mathcal{N}(0, I)$, where $I$ is the identity matrix (Doersch, 2016). The KL divergence loss (Rolinek et al., 2019) is given by:

$$\mathcal{L}_{\text{KL}} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{d}\frac{1}{2}\left(-1 - \log(\sigma_{i,j}^2) + \mu_{i,j}^2 + \sigma_{i,j}^2\right), \tag{2}$$

160 where $\mu_{i,j}$ and $\sigma_{i,j}$ are the mean and standard deviation for the $j$-th dimension of the latent variable $z_i$, and $d$ is the dimensionality of the latent space. This loss term ensures that the distribution of latent variables remains close to the desired prior distribution.

The total loss function for the VAE combines the reconstruction loss and the KL divergence loss:

$$\mathcal{L} = \mathcal{L}_{\text{rec}} + \mathcal{L}_{\text{KL}}. \tag{3}$$

165 By minimizing this total loss during training, the VAE learns to capture meaningful latent representations that can be used for fault detection and assessment. To enable backpropagation through the stochastic sampling process, the reparameterization trick (Doersch, 2016) is employed. Specifically, the latent variable $z$ is expressed as:

$$z_i = \mu_i + \sigma_i \cdot \epsilon, \quad \epsilon \sim \mathcal{N}(0, I), \tag{4}$$

where $\epsilon$ is a random variable sampled from a multivariate normal distribution, and $\mu_i$ and $\sigma_i$ are the mean and standard deviation
170 predicted by the encoder network. This reparameterization allows the gradients to flow through $\mu$ and $\sigma$ during optimization, enabling efficient training of the VAE. The training procedure is detailed in Algorithm 1.

https://doi.org/10.5194/wes-2025-123
Preprint. Discussion started: 6 August 2025
© Author(s) 2025. CC BY 4.0 License.

Open Access
WIND ENERGY SCIENCE
DISCUSSIONS

**Algorithm 1** Training procedure for the proposed fault detection method

**Input:** Dataset $D$, learning rate $\alpha$, batch size $B$, epochs $N_{\text{max}}$
**for** epoch $= 1$ to $N_{\text{max}}$ **do**
  **for** each batch $\{s_i\}_{i=1}^{B}$ from $D$ **do**
    Extract features $x_i$ from signal $s_i$
    $\mu_i, \sigma_i \leftarrow Q_\phi(x_i)$
    $z_i \leftarrow \mu_i + \sigma_i \odot \epsilon$, where $\epsilon \sim \mathcal{N}(0, I)$
    $\hat{x}_i \leftarrow P_\theta(z_i)$
    Compute $\mathcal{L} = \mathcal{L}_{\text{rec}} + \mathcal{L}_{\text{KL}}$
    Update $\theta, \phi$ using Adam optimizer
  **end for**
**end for**
**Return:** Trained parameters $\theta^*, \phi^*$

In this work, the input and output layers of the VAE each contain 280 neurons, corresponding to the 280 features derived from vibration signal analysis (20 time- and frequency-domain features from 14 channels). This ensures that the network can encode and reconstruct the complete set of features effectively. The encoder compresses the input data through three hidden layers with 128, 64, and 32 neurons, respectively, each followed by batch normalization and ReLU activation to ensure stable and efficient training. The latent space is represented by two fully connected layers: one for the mean (16 neurons) and one for the log variance (16 neurons), resulting in a latent representation size of 16. The decoder reconstructs the input features by symmetrically expanding the latent representation through layers with 32, 64, and 128 neurons, before mapping it back to the original 280-dimensional space in the output layer. Batch normalization and ReLU activations are also used in the decoder to maintain stability and improve performance.

### 3.4 Exponentially Weighted Moving Average

The exponentially weighted moving average is applied as the final step in monitoring the health of the wind turbine. This statistical tool is particularly effective in detecting small shifts or trends in data over time, making it ideal for fault detection in wind turbines (Sun et al., 2023; He et al., 2023). The EWMA smooths the reconstruction errors (REs) obtained from VAE, reducing noise and highlighting deviations indicative of potential faults.

The EWMA statistic, denoted as $Z_t$, is calculated recursively as:

$$Z_t = \lambda \cdot \text{RE}_t + (1 - \lambda) \cdot Z_{t-1}, \tag{5}$$

where $\text{RE}_t$ is the reconstruction error at the time $t$, and $\lambda$ is the smoothing parameter ($0 < \lambda < 1$), controlling the influence of recent errors versus historical trends. The smoothing parameter $\lambda$ controls the sensitivity of the EWMA. Smaller values of $\lambda$ assign more weight to historical data, reducing the impact of short-term fluctuations and producing a smoother curve. This increases stability but may delay the detection of abrupt changes. On the other hand, larger values of $\lambda$ give more weight to

https://doi.org/10.5194/wes-2025-123
Preprint. Discussion started: 6 August 2025
© Author(s) 2025. CC BY 4.0 License.

recent errors, making the chart more sensitive to sudden deviations but potentially more prone to false alarms caused by noise. In monitoring applications, the typical values for $\lambda$ are between 0.05 and 0.3 (He et al., 2023; Su et al., 2022; Liu et al., 2021). In this study, $\lambda$, set to 0.2, is selected as a compromise, providing a balance that allows the EWMA to respond to significant
195 anomalies while maintaining robustness against transient noise.

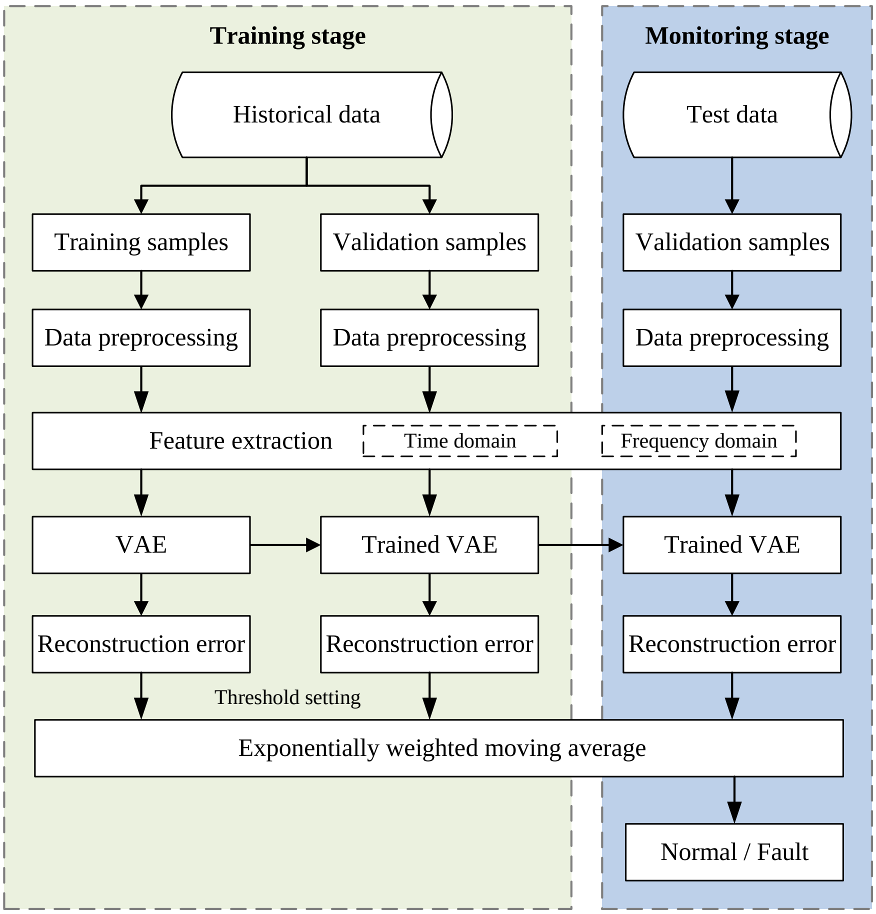


**Figure 2.** Flowchart of the HI construction approach.

### 3.5 Proposed HI Construction Method

The proposed health index construction method, as outlined in Figure 2, consists of a training stage and a monitoring stage, with the following detailed steps:

**Training stage:** In the training stage, historical vibration data collected during the normal operation of the wind turbine is
200 first preprocessed to remove outliers and trends, ensuring that the dataset accurately reflects the normal behavior of the turbine. The data is then divided into training and validation samples. Training samples are used to extract features and train the model,

https://doi.org/10.5194/wes-2025-123
Preprint. Discussion started: 6 August 2025
© Author(s) 2025. CC BY 4.0 License.
eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

while validation samples are used to set threshold limits to decide when a sample is diagnosed as faulty. Features are extracted from both the time and frequency domains to comprehensively characterize the signal's dynamics, resulting in 20 distinct attributes that capture critical aspects of the vibration signals. A VAE model is then trained on the features extracted from the training samples to learn the normal operating patterns of the turbine. The VAE encodes these features into a latent space and reconstructs them, minimizing the reconstruction error for normal data.

Once the VAE model is trained, the REs are computed for both the training and validation datasets. These REs are then smoothed using an EWMA, which reduces noise and highlights trends in the data. The threshold for anomaly detection is set as the maximum EWMA value observed in the combined training and validation datasets. This threshold serves as the baseline for distinguishing normal from abnormal conditions during monitoring. This approach is rooted in the assumption that the training and validation datasets represent normal and healthy turbine operations. Since these datasets are drawn exclusively from normal conditions, no alarms should be triggered within them. Thus, the maximum EWMA value from these datasets serves as a reliable benchmark or threshold: it represents the upper limit of what can be considered normal for healthy turbine operations.

**Monitoring stage** In the monitoring stage, real-time vibration data collected during turbine operations undergo the same preprocessing and feature extraction steps as in the training stage. The extracted features are then inputted into the trained VAE, which computes the reconstruction errors for each test sample. These reconstruction errors are fed into the EWMA, which tracks deviations from the normal operational baseline established during the training stage. The EWMA smooths the fluctuations in reconstruction errors, enabling the detection of gradual trends or abrupt changes. If the EWMA value exceeds the predefined threshold, the turbine condition is marked as abnormal, signaling the potential onset of a fault. In contrast, if the HI remains within the threshold, the turbine is classified as operating under normal conditions.

## 4 Experimental Validation

This section details the experimental validation of the proposed fault detection framework. It begins by describing the dataset and the specific fault scenarios used, followed by a summary of the data pre-processing steps, and finally, a thorough presentation and discussion of the monitoring results.

### 4.1 Experimental Dataset Description

The experimental dataset used is sourced from the Aventa AV-7 Research Wind Turbine, located in Taggenberg, which is managed by the ETH Zurich Department of Structural Health Monitoring (Chatzi et al., 2023). To comprehensively monitor the operational health of the wind turbine, vibration sensors are placed in critical locations, including the tower, the nacelle, the main shaft bearing, and the generator, as shown in Figure 3. These locations are chosen because they represent key components of the turbine where faults, such as structural instability, aerodynamic imbalances, or mechanical wear, are most likely to originate or propagate. By analyzing the vibration signals from these locations, the framework is able to capture a wide range of potential anomalies.

https://doi.org/10.5194/wes-2025-123
Preprint. Discussion started: 6 August 2025
© Author(s) 2025. CC BY 4.0 License.


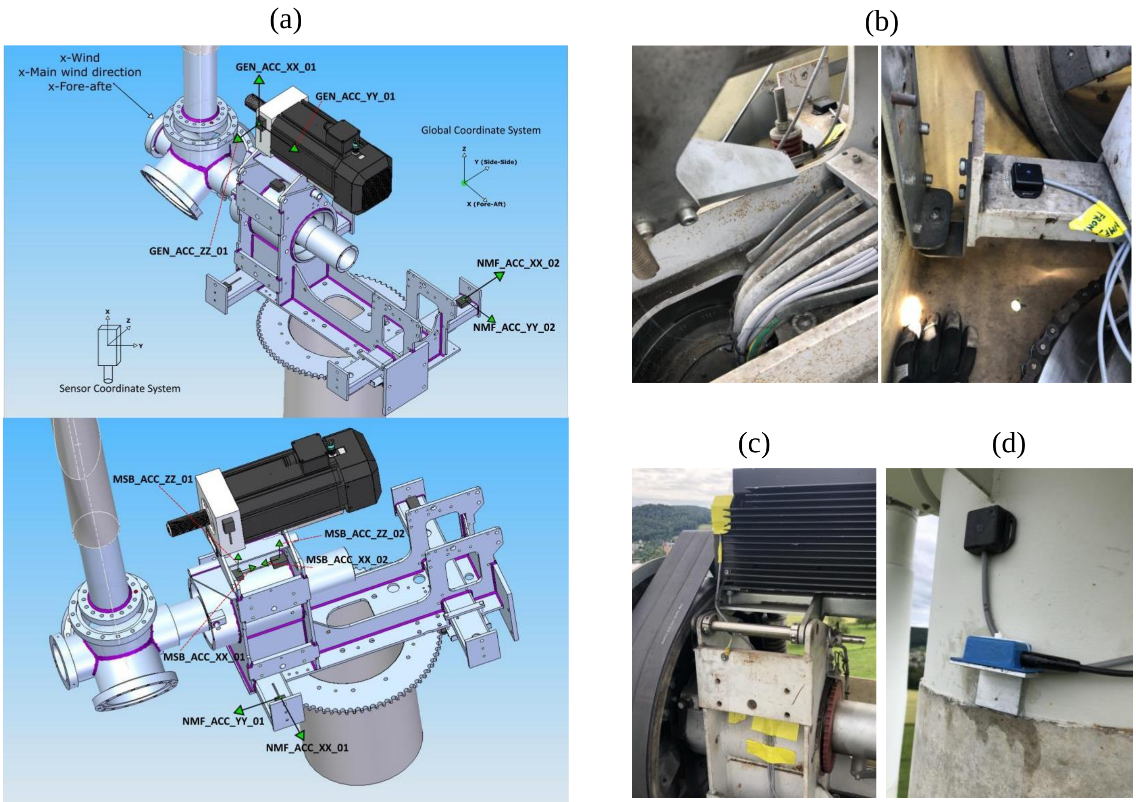


**Figure 3.** Sensor layout for wind turbine measurements (Chatzi et al., 2023): (a) diagram of the sensor placement on the nacelle; (b) accelerometers mounted on the nacelle main frame; (c) accelerometers installed on the bearing housing and generator; (d) accelerometers positioned on the tower top transition piece.

The analysis uses data from 14 specific vibration channels (Table 2) that are highly relevant to diagnosing the health of critical turbine components. The multi-location sensor placement ensures robust fault detection for various fault types. The complete operational history, including maintenance activities and fault periods, is summarized in the timeline in Figure 4. This study specifically focuses on three distinct fault scenarios, chosen to represent a diverse set of real-world challenges: a

**Table 2.** Channel names and corresponding locations of vibration signals used in the analysis.

| Location | Channel name | | | |
|---|---|---|---|---|
| Tower | L5_ACC_XX_01 | L5_ACC_YY_01 | L5_ACC_XX_02 | L5_ACC_YY_02 |
| Nacelle | NMF_ACC_YY_01 | NMF_ACC_XX_02 | NMF_ACC_YY_02 | |
| Main Shaft Bearing | MSH_ACC_XX_01 | MSH_ACC_ZZ_01 | MSB_ACC_XX_01 | MSB_ACC_ZZ_02 |
| Generator | GEN_ACC_XX_01 | GEN_ACC_YY_01 | GEN_ACC_ZZ_01 | |

https://doi.org/10.5194/wes-2025-123
Preprint. Discussion started: 6 August 2025
© Author(s) 2025. CC BY 4.0 License.

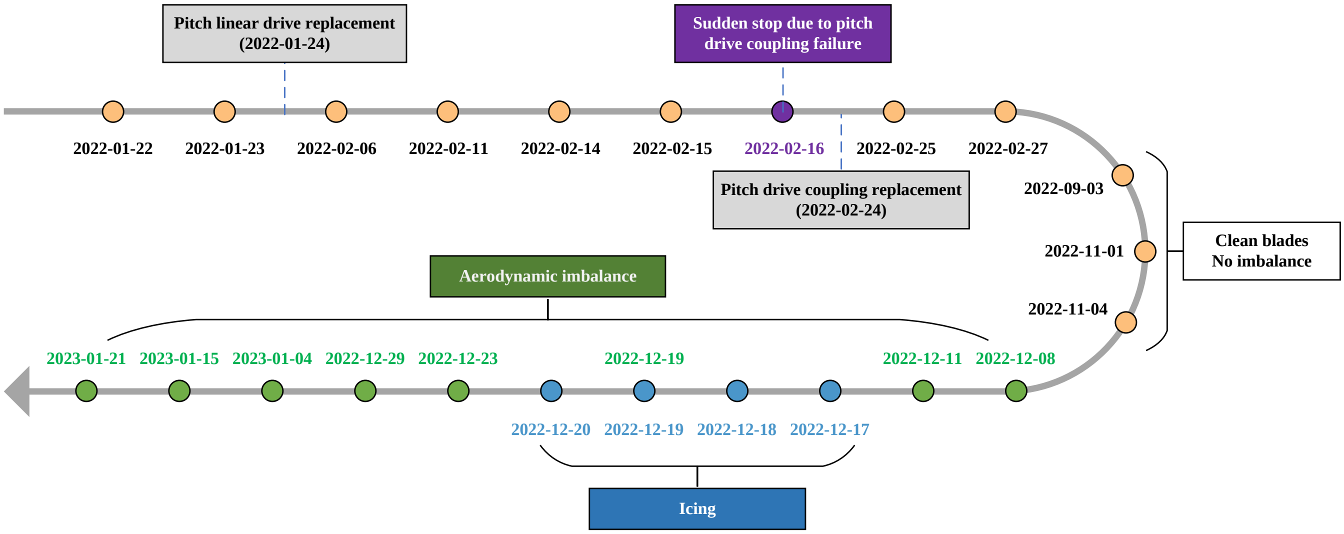


**Figure 4.** Timeline of failures and maintenance for Aventa AV-7 ETH Zurich research wind turbine.

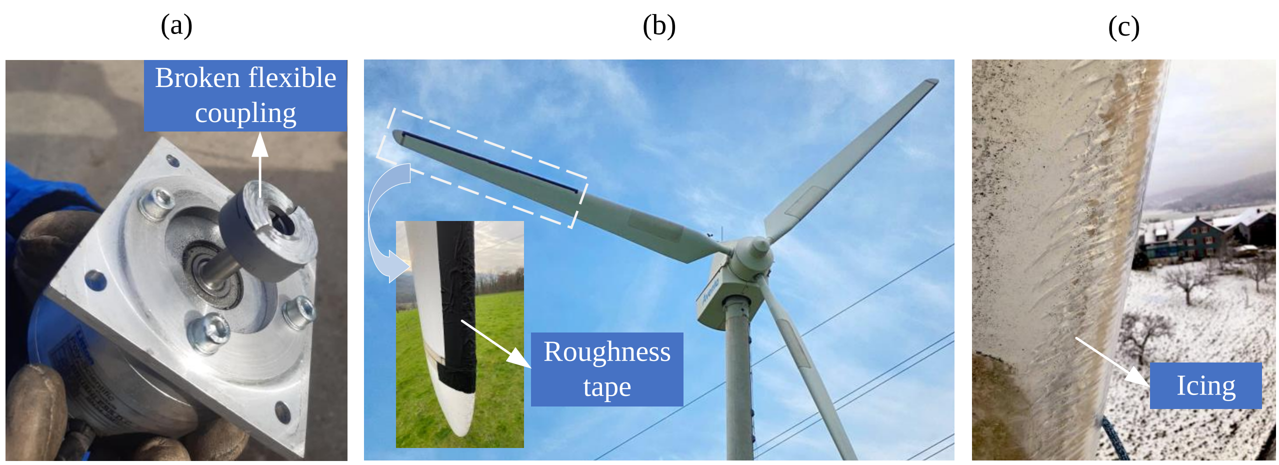


**Figure 5.** Illustration of wind turbine events (Chatzi et al., 2023): (a) failure of the pitch drive system caused by a broken flexible coupling; (b) aerodynamic imbalance due to roughness tape applied to the blade, simulating surface roughness effects; (c) icing on the blade.

mechanical breakdown, a simulated operational anomaly, and an environmental fault. These events, illustrated in Figure 5, include:

1. **Pitch Drive Failure**: A mechanical failure of a pitch drive coupling occurred on February 16, 2022, when a flexible coupling in the pitch drive system broke, as shown in Figure 5(a). This event led to a complete turbine stop, representing a critical component failure.

https://doi.org/10.5194/wes-2025-123
Preprint. Discussion started: 6 August 2025
© Author(s) 2025. CC BY 4.0 License.

2. **Aerodynamic Imbalance**: This condition was intentionally simulated from December 19, 2022, to January 15, 2023. As depicted in Figure 5(b), roughness tape was applied to the blade, simulating surface roughness effects that alter aerodynamic efficiency.

3. **Icing Events**: Naturally occurring blade icing was recorded during a period of cold weather from December 10 to December 19, 2022. This environmental fault, shown in Figure 5(c), affects both aerodynamic performance and rotational balance.

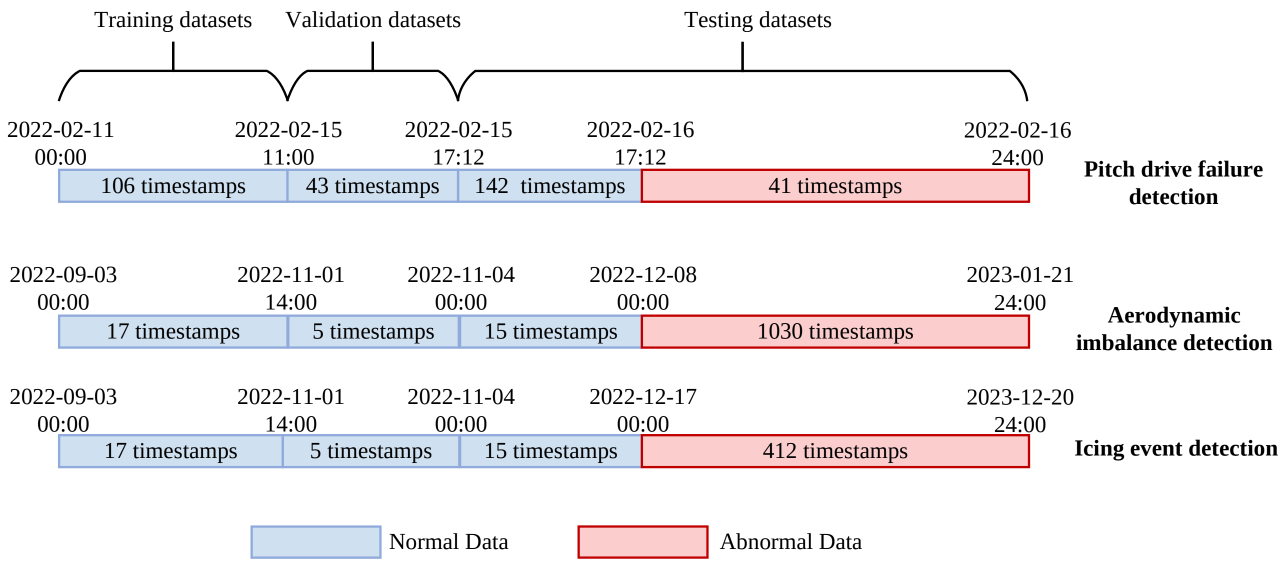


**Figure 6.** Division scheme of the training set, validation set, and the testing set.

**Table 3.** Number of timestamps and samples for training, validation, and testing datasets in three detection tasks.

| Task | Training datasets | Validation datasets | Test datasets |
|---|---|---|---|
| Pitch fault detection | 106 timestamps (6360 samples) | 43 timestamps (2580 samples) | 183 timestamps (10980 samples) |
| Aerodynamic imbalance detection | 17 timestamps (1020 samples) | 5 timestamps (300 samples) | 1045 timestamps (62700 samples) |
| Icing detection | 17 timestamps (1020 samples) | 5 timestamps (300 samples) | 427 timestamps (25620 samples) |

To prepare the data for the semi-supervised model, the dataset is partitioned into training, validation, and testing sets according to the operational timeline (Figure 6). A key aspect of this division is that the training and validation sets consist

https://doi.org/10.5194/wes-2025-123
Preprint. Discussion started: 6 August 2025
© Author(s) 2025. CC BY 4.0 License.


exclusively of data from confirmed healthy operational periods. This ensures that the VAE learns a true representation of the turbine's normal behavior. The test sets, conversely, contain both normal and faulty data, allowing for a robust evaluation of the framework's ability to distinguish between these states. The specific size of each dataset split is detailed in Table 3.

### 4.2 Data Pre-processing

The quality and reliability of the anomaly detection method heavily depend on proper data pre-processing. In this study, a comprehensive pipeline is implemented to ensure data quality and prepare raw vibration signals for subsequent feature extraction and model training. This pipeline includes outlier removal, data segmentation, detrending, and feature normalization.

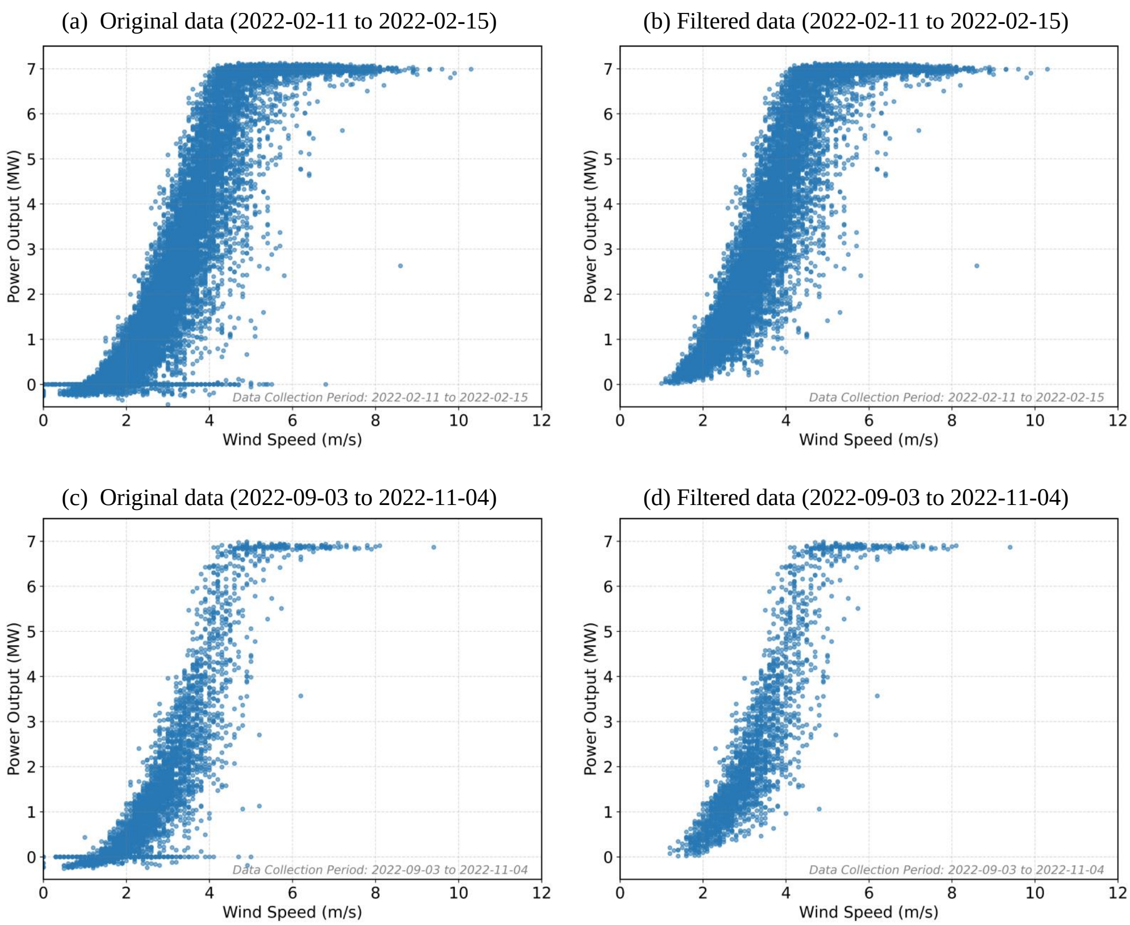


**Figure 7.** Comparison of power curves before and after outlier removal for two time periods.

https://doi.org/10.5194/wes-2025-123
Preprint. Discussion started: 6 August 2025
© Author(s) 2025. CC BY 4.0 License.

#### 4.2.1 Outlier Removal

To ensure that the training and validation datasets accurately represent true healthy operating conditions, anomalies in the power measurements (e.g., zero or negative values) are systematically identified and removed as outliers. Discarding these timestamps ensures that the dataset more accurately represents the normal operating conditions of the wind turbines. Figure 7 clearly illustrates the effectiveness of this step, comparing power curves before (Figure 7(a) and (c)) and after (Figure 7(b) and (d)) outlier removal for two distinct time periods.

#### 4.2.2 Data Segmentation

Each record or timestamp consists of approximately 10 minutes of vibration data sampled at 200 Hz. To process these data, it is divided into 60 non-overlapping sub-samples, with each sub-sample containing 2000 data points (corresponding to 10 seconds of operation). Timestamps with less than 10 minutes of data are excluded to ensure uniformity. For each 10-second sub-sample, the reconstruction error is calculated. Subsequently, the average of these 60 individual REs is computed to derive a single ensemble RE for the entire 10-minute record. This ensemble averaging strategy provides a more stable and robust representation of the turbine's operational state, effectively minimizing the influence of transient fluctuations on the health indicator. The data segmentation process is visually explained in Figure 8.

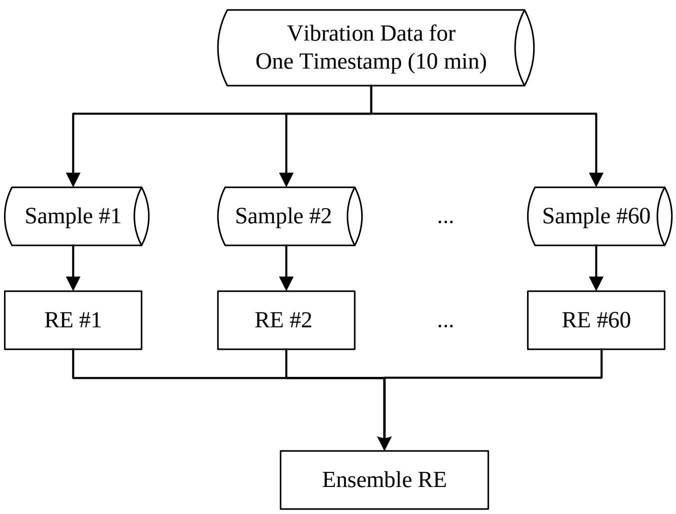


**Figure 8.** Illustration of the data segmentation process for vibration signals.

#### 4.2.3 Detrending

Unlike typical vibration signals that oscillate around zero, the experimental vibration data from the Aventa AV-7 exhibited exclusively positive values, likely due to the specific data acquisition settings. To counteract any inherent biases or drifts that could affect the analysis, detrending is applied to the vibration signals to ensure that the signals oscillate around zero.

https://doi.org/10.5194/wes-2025-123
Preprint. Discussion started: 6 August 2025
© Author(s) 2025. CC BY 4.0 License.

Specifically, in this work, detrending is achieved by subtracting the mean value of each sample from its own data points. This ensures that each sample is centered around zero, without any biases introduced by the data acquisition settings.

#### 4.2.4 Feature Normalization

Following extraction of features in the time and frequency domains from detrended vibration signals, min-max scaling is applied to normalize these features. This standardization step is critical for machine learning models, as it brings all features to a comparable scale, preventing features with larger numerical ranges from dominating the learning process. Crucially, the maximum and minimum values for scaling are computed exclusively from the training dataset, ensuring data consistency and standardization throughout model training while preventing information leakage from the test set.

### 4.3 Experimental Results

The performance of the proposed framework is systematically evaluated in three distinct fault scenarios, showcasing its robust capabilities to detect various types of anomalies. Each type of fault, including pitch drive failure, aerodynamic imbalance, and icing event, presents unique challenges in monitoring wind turbine health, thereby providing a comprehensive validation of the framework.

To demonstrate the effectiveness of the proposed approach, comparative analysis is conducted with several baseline methods using the same 14-channel vibration signals. The baseline methods include standard autoencoder (AE), denoising autoencoder (DAE) (Chen et al., 2020), deep support vector data description (Deep SVDD) (Peng et al., 2025), one-class support vector machine (OC-SVM) (Ghiasi et al., 2024), and a standard variational autoencoder (VAE). The configurations are as follows:

- To directly evaluate the impact of the input type, AE, DAE, and Deep SVDD are tested with both feature-based (F) and raw signal (R) inputs.
- The OC-SVM method, which conventionally operates on a feature space, is applied directly to the engineered features, referred to as OC-SVM$_{\mathrm{F}}$.
- To create a direct ablation baseline for our framework, a standard VAE is applied to the raw signal (VAE$_{\mathrm{R}}$).

For fair comparison, all baseline methods adopt identical network architectures and training procedures. The encoder compresses input data through three hidden layers with 128, 64, and 32 neurons, while the decoder symmetrically reconstructs the input through layers with 32, 64, and 128 neurons. Batch normalization and ReLU activations are applied in both encoder and decoder to maintain stability and improve performance. All models are implemented using PyTorch and optimized using the Adam optimizer with a learning rate of 0.001, batch size of 256, and maximum 100 epochs on an NVIDIA 4060 GPU.

https://doi.org/10.5194/wes-2025-123
Preprint. Discussion started: 6 August 2025
© Author(s) 2025. CC BY 4.0 License.

#### 4.3.1 Case I: Pitch Drive Failure

Figure 9 illustrates the result of the pitch drive failure monitoring using the proposed method. The green dots represent the HI during the training stage, while the blue dots show the HI values from the validation stage used to set the threshold (horizontal red dashed line). The orange dots depict the HI during the monitoring (testing) stage including both normal and faulty data.

The framework demonstrates exceptional early warning capability for this progressive fault. As shown in Figure 9, an alarm is triggered at 14:32, providing a 2.5-hour lead time before the actual turbine shutdown at 17:12. This substantial lead time enables proactive maintenance planning and minimizes unexpected downtime.

The performance of baseline methods is detailed in Figure 10. The methods with feature-based input (F) demonstrate superior early warning capabilities. Specifically, $AE_F$, $DAE_F$, Deep $SVDD_F$, and OC-$SVM_F$ all successfully provide early warnings approximately 2.5 hours before the actual turbine shutdown with HI values clearly crossing the threshold well in advance of fault occurrence. In contrast, the methods with raw signal input (R) exhibit significantly inferior performance patterns. $AE_R$ and $DAE_R$ fail to detect the fault entirely. Deep $SVDD_R$ and $VAE_R$ provide fault indication only at the moment of fault occurrence without early warning capability, while $VAE_R$ suffers from frequent false alarms during normal operation periods.

This comparison clearly demonstrates the superior performance of methods using feature-based input over those using raw signal input for complex fault detection tasks, highlighting the critical importance of proper feature engineering in achieving reliable early warning capabilities.

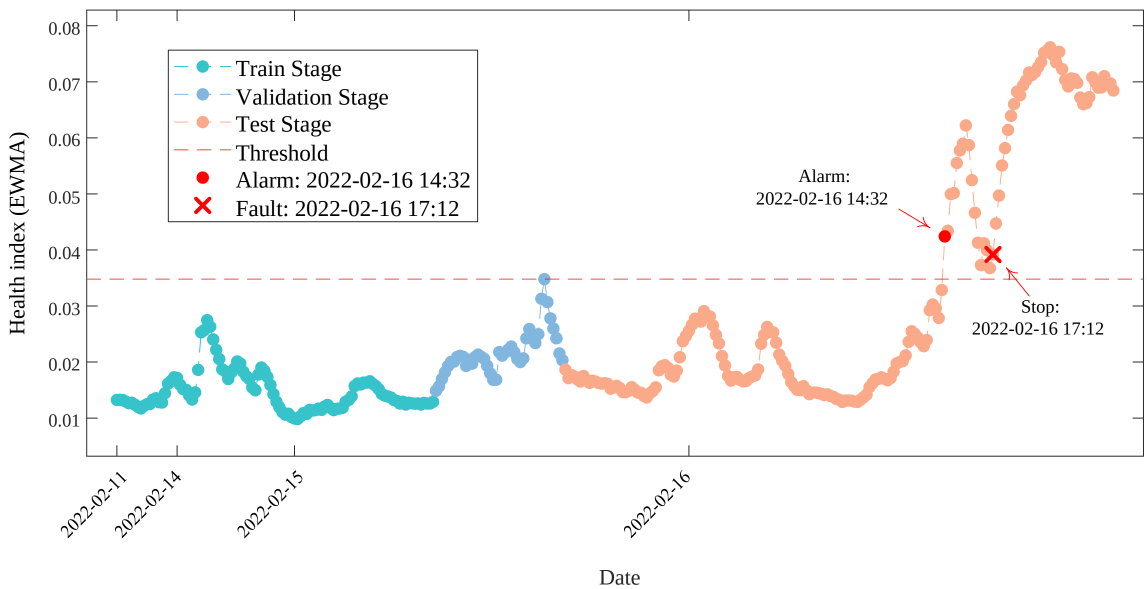


**Figure 9.** Monitoring result for wind turbine pitch drive failure using the proposed method.

https://doi.org/10.5194/wes-2025-123
Preprint. Discussion started: 6 August 2025
© Author(s) 2025. CC BY 4.0 License.
Open Access WIND ENERGY SCIENCE DISCUSSIONS

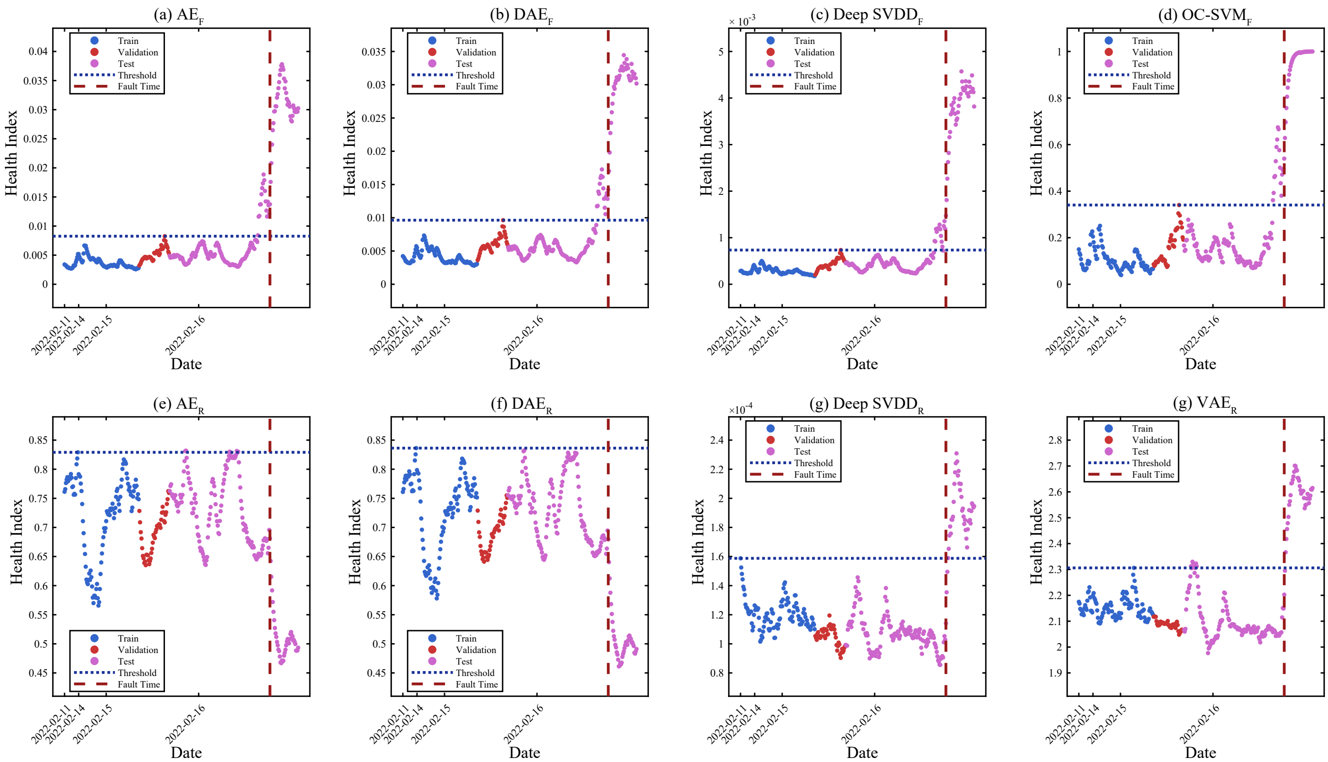


**Figure 10.** Health index comparison of baseline methods during Case I pitch drive failure detection. The figure shows (a-d) methods with feature-based input and (e-h) methods with raw signal input. The red dashed line indicates the fault occurrence time, while the blue dotted line represents the detection threshold.

### 4.3.2 Case II: Aerodynamic Imbalance

Figure 11 illustrates the monitoring results for aerodynamic imbalance detection using the proposed method. The red dashed threshold line represents the decision boundary for imbalance detection. On December 8, 2022, an imbalance event is accurately detected, as indicated by the spike in the health index that crosses the threshold line, triggering an immediate alarm. The proposed method achieves precise detection with no false alarms during normal operation periods and no missed detections during fault occurrence, demonstrating excellent discrimination capability between normal and faulty conditions.

The performance of the baseline methods, detailed in Figure 12, varies significantly in terms of reliability, even though all feature-based methods successfully identify the imbalance event. Specifically, both $AE_F$ and OC-$SVM_F$ are compromised by false alarms during normal operational phases. In contrast, Deep $SVDD_F$ provides a stable baseline free of false alarms but suffers from some missed detections during the fault period. Among all baselines, $DAE_F$ demonstrates exceptional performance in this case, achieving a perfect detection record with metrics identical to the proposed method. The methods using raw signal

https://doi.org/10.5194/wes-2025-123
Preprint. Discussion started: 6 August 2025
© Author(s) 2025. CC BY 4.0 License.


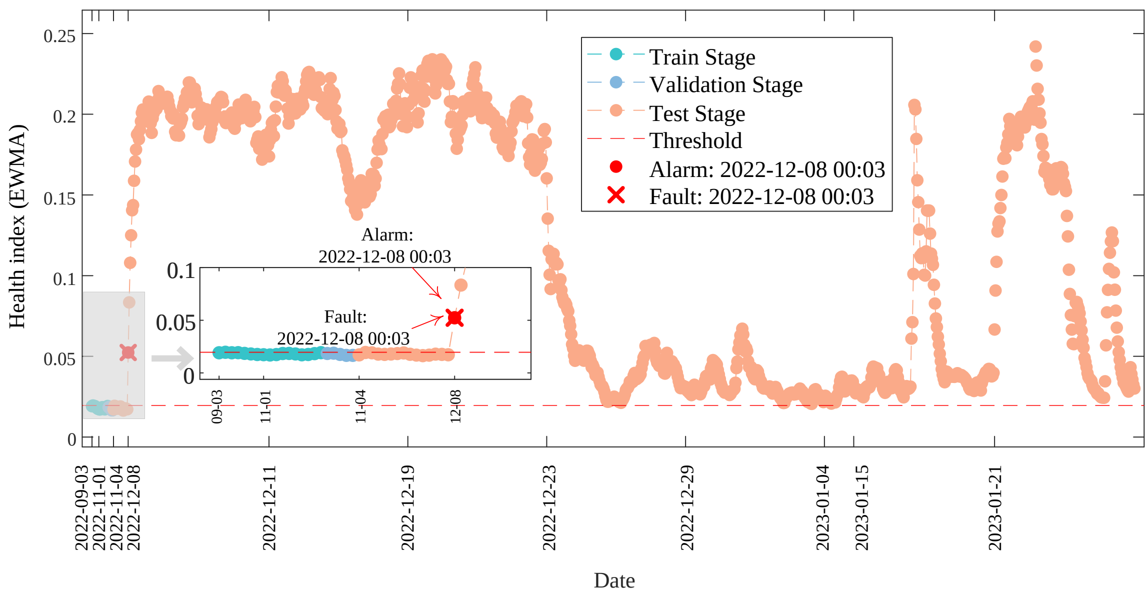


**Figure 11.** Monitoring result for wind turbine aerodynamic imbalance using the proposed method.

input (R variants) proved largely ineffective, suffering from high rates of both false alarms and missed detections. This again highlights the proposed framework's consistent and reliable performance across different conditions.

#### 4.3.3 Case III: Icing Event

The monitoring process for detecting icing events using the proposed method is illustrated in Figure 13. On December 17, 2022, a significant spike in the EWMA value crosses the red dashed threshold line, triggering an alarm. This rapid increase in the HI value provides a clear indication of the fault's occurrence. The results demonstrate that the proposed framework accurately identifies the icing event without any false alarms or missed detections, as evidenced by the clear distinction between normal and faulty conditions in the figure.

The baseline comparison in Figure 14 shows that while most methods could detect the icing event, their reliability varied significantly. All feature-based methods (F variants) once again provided clear and stable fault signatures. The most notable deficiency is observed in the raw-signal variants of AE and DAE, where both $AE_R$ and $DAE_R$ suffer from high baseline variability, leading to numerous false alarms during periods of normal operation. The other baseline methods are able to identify the fault without this significant false alarm issue. This result further underscores the superiority of the proposed framework, which delivers accurate fault detection while ensuring a stable and reliable baseline.

#### 4.3.4 Performance metrics

To quantitatively evaluate and compare the fault detection performance of the different HI construction methods, Table 4 presents a comprehensive quantitative comparison of different methods for aerodynamic imbalance (Case II) and icing events

https://doi.org/10.5194/wes-2025-123
Preprint. Discussion started: 6 August 2025
© Author(s) 2025. CC BY 4.0 License.


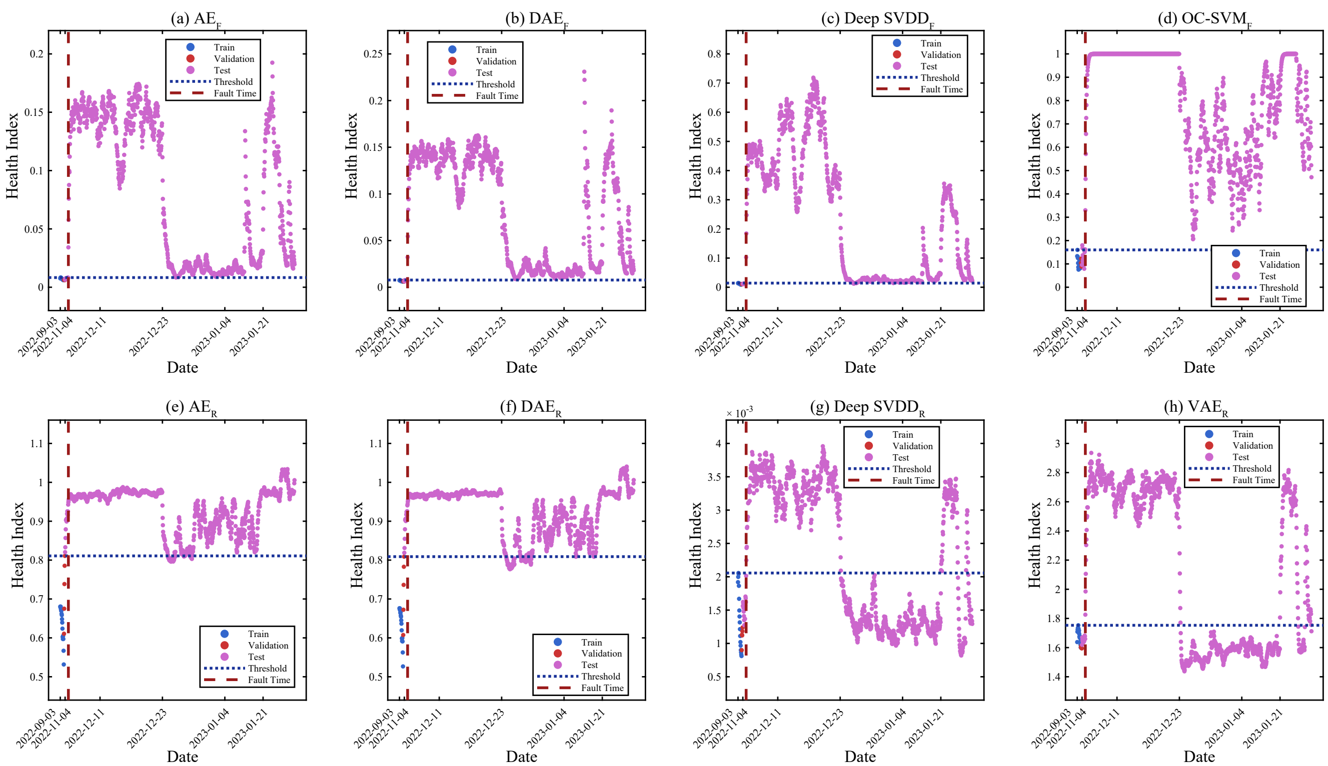


**Figure 12.** Health index comparison of baseline methods during Case II aerodynamic imbalance detection. The figure shows (a-d) methods with feature-based input and (e-h) methods with raw signal input. The red dashed line indicates the fault occurrence time, while the blue dotted line represents the detection threshold.

(Case III). These two cases provide clearly defined fault labels based on operational records, enabling precise evaluation of false alarm rates and missed alarm rates alongside conventional accuracy metrics. Here, Case I (pitch drive failure) is excluded from this quantitative analysis as it is primarily intended for assessing the early warning capabilities of the models due to its progressive failure characteristics.

As shown in Table 4, the proposed framework consistently demonstrates excellent performance, achieving 100% accuracy with zero false alarm and missed alarm rates in both test scenarios. While the DAE$_F$ baseline also achieved a perfect score in Case II, our method is distinguished by its superior consistency across diverse fault types. This robust and reliable performance validates the effectiveness of our integrated approach.

Furthermore, the quantitative results confirm the superiority of methods using feature-based input over those using raw signal input under identical network architectures. This contrast highlights that well-engineered features can effectively reduce dimensionality and filter noise, thereby accentuating fault-related information crucial for robust detection from complex multi-channel vibration signals.

https://doi.org/10.5194/wes-2025-123
Preprint. Discussion started: 6 August 2025
© Author(s) 2025. CC BY 4.0 License.

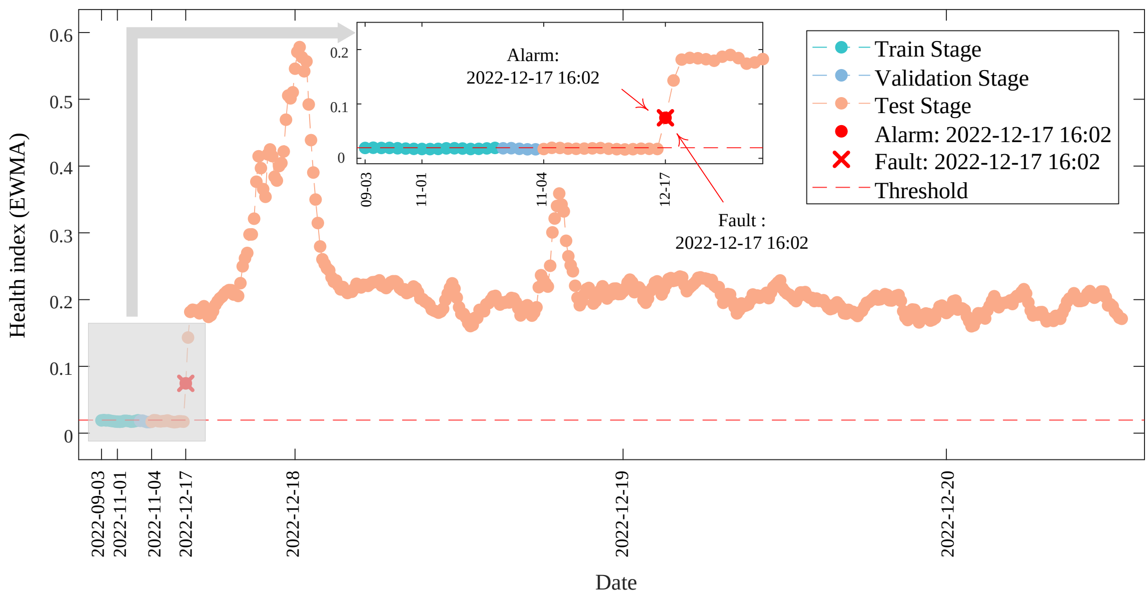


**Figure 13.** Monitoring result for wind turbine icing event using the proposed method.

In summary, the comprehensive evaluation validates that by synergizing effective feature engineering with a VAE-based
360 detector and EWMA smoothing, our framework offers a highly reliable and robust solution with strong potential for practical deployment.

**Table 4.** Performance comparison of different methods on imbalance and icing events.

| | Case II (Aerodynamic imbalance) | | | Case III (Icing events) | | |
|---|---|---|---|---|---|---|
| Method | Accuracy | False alarm rate | Missed alarm rate | Accuracy | False alarm rate | Missed alarm rate |
| $AE_R$ | 0.9474 | 1.0000 | 0.0388 | 0.9649 | 1.0000 | 0.0000 |
| $DAE_R$ | 0.9043 | 1.0000 | 0.0825 | 0.9649 | 1.0000 | 0.0000 |
| Deep $SVDD_R$ | 0.5100 | 0.0000 | 0.4971 | 0.9977 | 0.0000 | 0.0024 |
| $VAE_R$ | 0.5368 | 0.0000 | 0.4699 | 1.0000 | 0.0000 | 0.0000 |
| $AE_F$ | 0.9990 | 0.0667 | 0.0000 | 0.9977 | 0.0667 | 0.0000 |
| $DAE_F$ | 1.0000 | 0.0000 | 0.0000 | 0.9977 | 0.0667 | 0.0000 |
| Deep $SVDD_F$ | 0.9923 | 0.0000 | 0.0078 | 1.0000 | 0.0000 | 0.0000 |
| OC-$SVM_F$ | 0.9971 | 0.2000 | 0.0000 | 0.9930 | 0.2000 | 0.0000 |
| Proposed | **1.0000** | **0.0000** | **0.0000** | **1.0000** | **0.0000** | **0.0000** |

https://doi.org/10.5194/wes-2025-123
Preprint. Discussion started: 6 August 2025
© Author(s) 2025. CC BY 4.0 License.



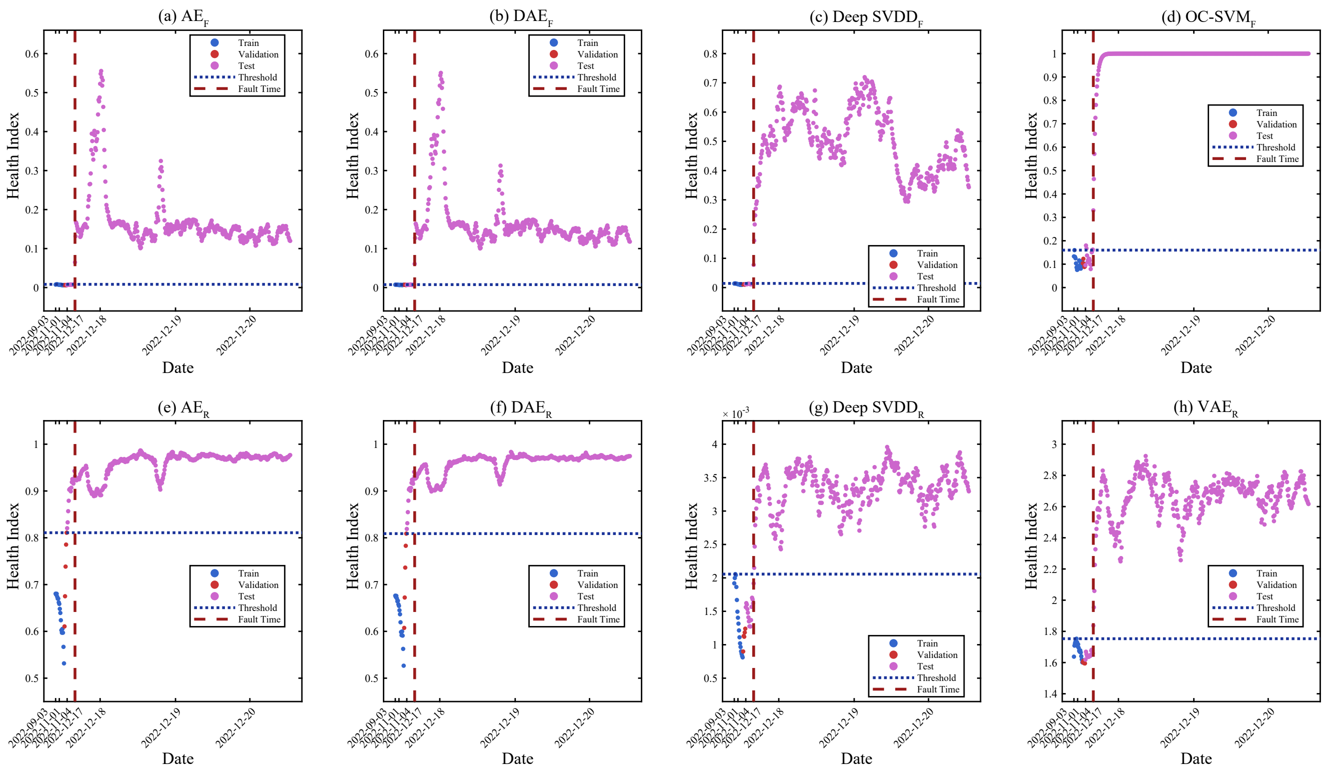


**Figure 14.** Health index comparison of baseline methods during Case III icing event detection. The figure shows (a-d) methods with feature-based input and (e-h) methods with raw signal input. The red dashed line indicates the fault occurrence time, while the blue dotted line represents the detection threshold.

### 4.4 Impact of $\lambda$ in EWMA

In addition, the impact of the smoothing parameter $\lambda$ in the EWMA method on the fault detection performance is analyzed. Table 5 summarizes the performance metrics (accuracy, recall, precision, and $F_1$ score) for different values of the parameter $\lambda$ in the context of detection of aerodynamic imbalances.

Higher values of $\lambda$ (e.g., 0.45 and 0.4), assign greater weight to recent errors, making the EWMA chart more sensitive to abrupt deviations. However, this increased sensitivity can result in a higher likelihood of false alarms caused by noise. As $\lambda$ decreases to 0.35 and 0.3, the performance improves, achieving near-perfect results with minimal false positives or missed anomalies. At $\lambda = 0.25$, the model achieves perfect scores in all metrics, with the accuracy, recall, precision, and $F_1$ scores reaching 100%. This trend continues for even smaller values of $\lambda$ (e.g. 0.2 and 0.15), where the model consistently maintains perfect detection performance. The results show that lower values of $\lambda$ result in a progressively smoother EWMA curve, reducing false alarms while maintaining accurate and early detection of anomalies. Based on this analysis, a value of $\lambda = 0.2$

https://doi.org/10.5194/wes-2025-123
Preprint. Discussion started: 6 August 2025
© Author(s) 2025. CC BY 4.0 License.

**Table 5.** The performance metrics (accuracy, recall, precision, and $F_1$ score) for different values of $\lambda$ in the EWMA method for aerodynamic imbalance.

| $\lambda$ | Accuracy | Recall | Precision | $F_1$ Score |
|---|---|---|---|---|
| 0.45 | 98.76% | 98.83% | 99.90% | 99.37% |
| 0.40 | 98.95% | 99.13% | 99.80% | 99.46% |
| 0.35 | 99.23% | 99.42% | 99.81% | 99.61% |
| 0.30 | 99.90% | 100.00% | 99.90% | 99.95% |
| 0.25 | 100.00% | 100.00% | 100.00% | 100.00% |
| 0.20 | 100.00% | 100.00% | 100.00% | 100.00% |
| 0.15 | 100.00% | 100.00% | 100.00% | 100.00% |

is chosen for the proposed framework, ensuring reliable anomaly detection with high and consistent performance in all metrics evaluated.

### 4.5 Computational Efficiency

The computational efficiency of the proposed framework is assessed to ensure its suitability for real-time wind turbine monitoring and fault detection, as shown in Table 6. All models are implemented using PyTorch, and training and inference tasks are executed on hardware with the following specifications: GPU: NVIDIA GeForce RTX 4060, CPU: Intel Core i7-13700K, Memory: 32 GB RAM. Feature extraction, which involves computing 20 time- and frequency-domain features across 14 vibration signal channels, requires an average of 2.53 seconds per 10-minute dataset. Model training times vary by task, reflecting the size and complexity of the dataset. The inference process, which includes feature extraction, model loading, feature reconstruction, and HI computation for anomaly detection, required an average of 2.56 seconds per 10-minute dataset. These computational times demonstrate the practical viability of the framework for continuous, near-real-time monitoring applications in wind turbine operations.

### 4.6 Discussion

This section provides a more profound insight into the framework's ability to interpret and differentiate between various fault types based on the HI values, particularly focusing on icing events and aerodynamic imbalances. Figure 15 graphically illustrates the HI values alongside the relevant environmental data.

Rotor icing not only alters the aerodynamic properties of the blades but also introduces mass imbalance, leading to pronounced deviations in vibration signals. These combined effects typically result in higher reconstruction errors and higher HI values. In contrast, the aerodynamic imbalance caused by the roughness tape is primarily localized, with less severe effects on the rotor dynamics, often producing lower HI values. As shown in Figures 15 (a) and (b), HI values during icing events are

https://doi.org/10.5194/wes-2025-123
Preprint. Discussion started: 6 August 2025
© Author(s) 2025. CC BY 4.0 License.

**Table 6.** Computational efficiency for feature extraction, model training, and inference.

| Process | Task | Time cost | Notes |
|---|---|---|---|
| Feature Extraction | All Tasks | 2.53 seconds per 10-minute dataset | Computation of 20 time-domain and frequency-domain features for 14 channels. |
| Model Training | Pitch Drive Failure | 6.81 seconds | Trained for 100 epochs on the given feature dataset. |
| | Aerodynamic Imbalance | 1.27 seconds | |
| | Icing Detection | 1.27 seconds | |
| Inference | All Tasks | 2.56 seconds per 10-minute dataset | Includes feature extraction, model loading, feature reconstruction, and HI computation. |

generally higher than those observed during aerodynamic imbalance periods, except for the period from December 8 to 11, 2022.

In particular, as can be seen in Figure 15 (b), during the period from December 8 to 11, 2022, highlighted in the green region, the framework detected HI values comparable to those observed during the confirmed icing event from December 17 to 20, 2022. Analysis of environmental data, including low temperatures and high humidity (Figure 15 (c) and (d)), suggests that the conditions were conducive to icing formation. This raises the hypothesis that icing may have co-occurred with aerodynamic imbalance during this period. Similarly, there are potential indications of icing events on January 15 and 21, 2023, based on similar environmental conditions.

This ability to identify periods with high HI values correlating with conducive environmental conditions, even without direct fault labels, underscores the framework's potential for proactive and insightful operational diagnostics. If these potential icing events were to be confirmed, a simple threshold on the HI values could effectively differentiate between icing and aerodynamic imbalance, based on their distinct HI magnitudes. The results from this analysis highlight the framework's advanced potential not only for detecting anomalies but also for contributing to the differentiation of fault types based on their characteristic HI signatures.

## 5 Conclusion

This paper presented a semi-supervised fault detection framework that integrates multi-channel vibration analysis with deep learning. The core of the methodology is to build a robust model of a wind turbine's normal operational state using only healthy data. This is achieved by training a variational autoencoder on a comprehensive set of time-domain and frequency-domain features extracted from vibration signals. An exponentially weighted moving average is then applied to the VAE's reconstruction error to create a stable and reliable health index, minimizing the risk of false alarms. The framework's effectiveness is validated on the ETH Zurich research turbine dataset, covering three distinct fault types. The experimental results demonstrate a high

https://doi.org/10.5194/wes-2025-123
Preprint. Discussion started: 6 August 2025
© Author(s) 2025. CC BY 4.0 License.

level of performance. Notably, the framework provided a 2.5-hour early warning for a critical pitch drive failure and achieved perfect (100%) accuracy in detecting both simulated aerodynamic imbalances and naturally occurring icing events. The experimental results validate the framework's effectiveness in providing timely and accurate fault detection, offering sufficient lead time for proactive maintenance. These findings confirm that the approach provides timely and accurate fault detection, a capability crucial for enhancing operational reliability and minimizing downtime.

While the framework is highly effective for wind turbine fault detection, it currently does not perform detailed diagnosis. Future work will therefore focus on extending the methodology to include fault localization and fault identification by analyzing the signatures in feature-level and channel-level reconstruction errors.

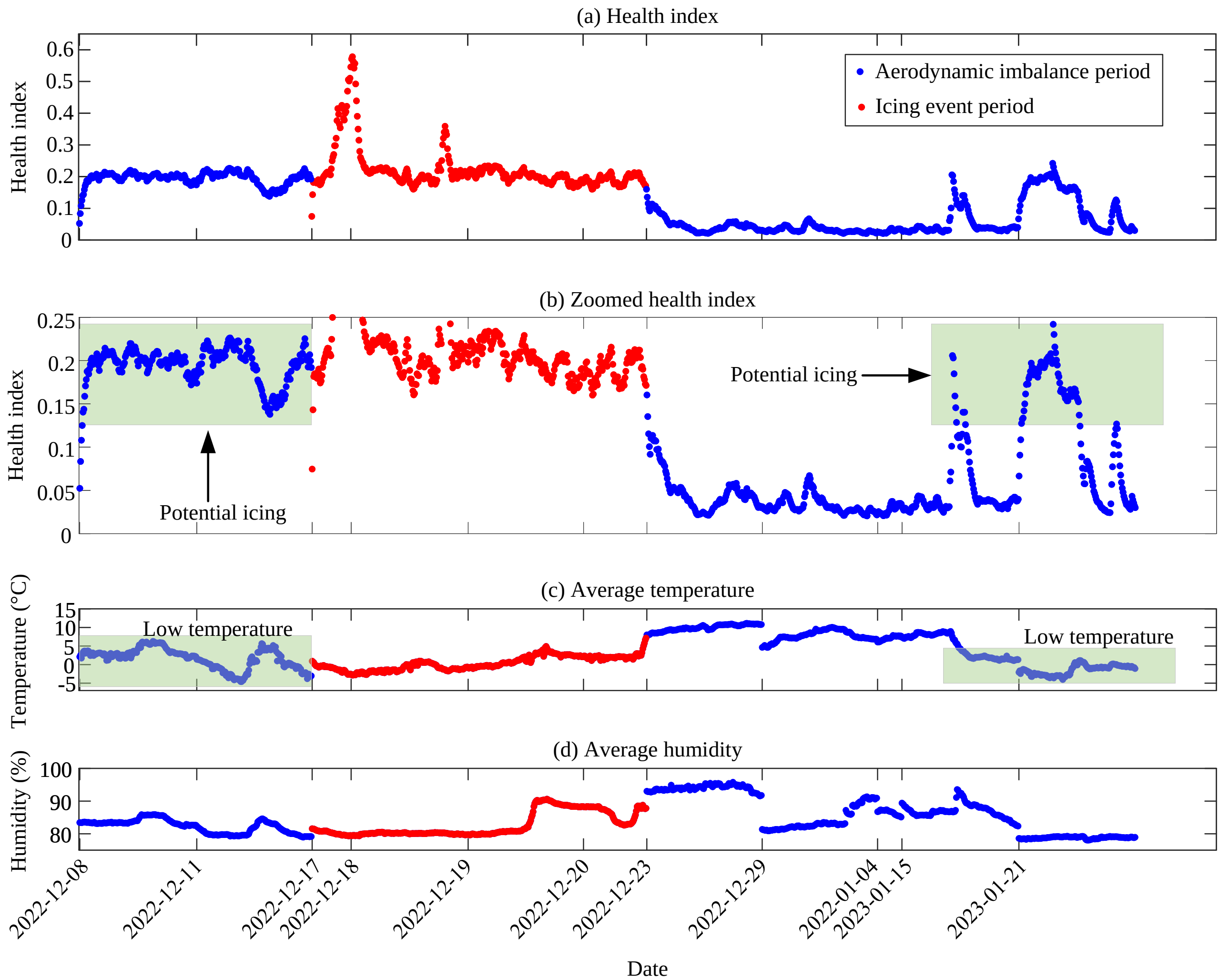


**Figure 15.** (a) Comparison of HI values for icing events and aerodynamic imbalance periods, (b) zoomed HI values, (c) temperature data trends, and (d) humidity data trends during the monitoring period.

https://doi.org/10.5194/wes-2025-123
Preprint. Discussion started: 6 August 2025
© Author(s) 2025. CC BY 4.0 License.

*Code availability.* The source code is publicly available at https://github.com/shun-wang1/wedowind-challenge-ASCE-EMI/tree/main.

*Data availability.* The challenge dataset was provided through Zenodo https://doi.org/10.5281/zenodo.8229750.

*Author contributions.* SW contributed to the paper by performing data analysis, developing methodology, and preparing the original draft of the manuscript. YV and FP contributed to the paper by conducting formal analysis, securing funding for the project, and reviewing and editing the manuscript.

*Competing interests.* Yolanda Vidal is a member of the editorial board of Wind Energy Science.

*Acknowledgements.* This work is partially funded by: $(i)$ grant PID2021-122132OB-C21 funded by MCIN/AEI/10.13039/501100011033 and by "ERDF A way of making Europe", by the "European Union"; $(ii)$ grant TED2021-129512B-I00 funded by MCIN/AEI/10.13039/501100011033 and by the "European Union NextGenerationEU/PRTR"; $(iii)$ grant 2021-SGR-01044 funded by the Generalitat de Catalunya.

https://doi.org/10.5194/wes-2025-123
Preprint. Discussion started: 6 August 2025
© Author(s) 2025. CC BY 4.0 License.

## References

Abid, A., Khan, M. T., and Iqbal, J.: A review on fault detection and diagnosis techniques: basics and beyond, Artificial Intelligence Review, 54, 3639–3664, 2021.

Ashkarkalaei, M., Ghiasi, R., Pakrashi, V., and Malekjafarian, A.: Feature selection for unsupervised defect detection of a wind turbine blade considering operational and environmental conditions, Mechanical Systems and Signal Processing, 230, 112 568, 2025.

Badihi, H., Zhang, Y., Jiang, B., Pillay, P., and Rakheja, S.: A comprehensive review on signal-based and model-based condition monitoring of wind turbines: Fault diagnosis and lifetime prognosis, Proceedings of the IEEE, 110, 754–806, 2022.

Bilendo, F., Lu, N., Badihi, H., Meyer, A., Cali, Ü., and Cambron, P.: Multitarget normal behavior model based on heterogeneous stacked regressions and change-point detection for wind turbine condition monitoring, IEEE Transactions on Industrial Informatics, 20, 5171–5181, 2023.

Castellani, F., Natili, F., Astolfi, D., and Vidal, Y.: Wind turbine gearbox condition monitoring through the sequential analysis of industrial SCADA and vibration data, Energy Reports, 12, 750–761, 2024.

Chatzi, E., Abdallah, I., Hofsäß, M., Bischoff, O., Barber, S., and Marykovskiy, Y.: Aventa AV-7 ETH Zurich Research Wind Turbine SCADA and High-Frequency Structural Health Monitoring (SHM) Data, https://doi.org/10.5281/zenodo.8229750, 2023.

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, 2021.

Chen, J., Li, J., Chen, W., Wang, Y., and Jiang, T.: Anomaly detection for wind turbines based on the reconstruction of condition parameters using stacked denoising autoencoders, Renewable Energy, 147, 1469–1480, 2020.

Chen, W., Zhou, H., Cheng, L., and Xia, M.: Condition Monitoring and Anomaly Detection of Wind Turbines using Temporal Convolutional Informer and Robust Dynamic Mahalanobis Mahalanobis Distance, IEEE Transactions on Instrumentation and Measurement, 2023.

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, 2023.

Council, G. W. E.: Global Wind Report 2025, https://www.gwec.net/reports/globalwindreport/, 2025.

Cuesta, J., Leturiondo, U., Vidal, Y., and Pozo, F.: A review of prognostics and health management techniques in wind energy, Reliability Engineering & System Safety, p. 111004, 2025.

Dey, S., Pisu, P., and Ayalew, B.: A comparative study of three fault diagnosis schemes for wind turbines, IEEE Transactions on Control Systems Technology, 23, 1853–1868, 2015.

Dibaj, A., Gao, Z., and Nejad, A. R.: Fault detection of offshore wind turbine drivetrains in different environmental conditions through optimal selection of vibration measurements, Renewable Energy, 203, 161–176, 2023.

Ding, P., Jia, M., and Yan, X.: Stationary subspaces-vector autoregressive with exogenous terms methodology for degradation trend estimation of rolling and slewing bearings, Mechanical Systems and Signal Processing, 150, 107 293, 2021.

Doersch, C.: Tutorial on variational autoencoders, arXiv preprint arXiv:1606.05908, 2016.

Encalada-Dávila, Á., Moyón, L., Tutivén, C., Puruncajas, B., and Vidal, Y.: Early fault detection in the main bearing of wind turbines based on Gated Recurrent Unit (GRU) neural networks and SCADA data, IEEE/ASME Transactions on Mechatronics, 27, 5583–5593, 2022.

Ghiasi, R., Khan, M. A., Sorrentino, D., Diaine, C., and Malekjafarian, A.: An unsupervised anomaly detection framework for onboard monitoring of railway track geometrical defects using one-class support vector machine, Engineering Applications of Artificial Intelligence, 133, 108 167, 2024.

https://doi.org/10.5194/wes-2025-123
Preprint. Discussion started: 6 August 2025
© Author(s) 2025. CC BY 4.0 License.

Habibi, H., Howard, I., and Simani, S.: Reliability improvement of wind turbine power generation using model-based fault detection and fault tolerant control: A review, Renewable energy, 135, 877–896, 2019.

He, Y., Liu, J., Wu, S., and Wang, X.: Condition monitoring and fault detection of wind turbine driveline with the implementation of deep residual long short-term memory network, IEEE Sensors Journal, 23, 13 360–13 376, 2023.

Irena, I.: Renewable power generation costs in 2017. Report, International Renewable Energy Agency, Abu Dhabi, 2018.

Jiang, G., Xie, P., He, H., and Yan, J.: Wind turbine fault detection using a denoising autoencoder with temporal information, IEEE/Asme transactions on mechatronics, 23, 89–100, 2017.

Jiang, G., He, H., Yan, J., and Xie, P.: Multiscale convolutional neural networks for fault diagnosis of wind turbine gearbox, IEEE Transactions on Industrial Electronics, 66, 3196–3207, 2018.

Jiang, Y., Tang, B., Qin, Y., and Liu, W.: Feature extraction method of wind turbine based on adaptive Morlet wavelet and SVD, Renewable energy, 36, 2146–2153, 2011.

Jin, X., Lv, S., Kong, Z., Yang, H., Zhang, Y., Guo, Y., and Xu, Z.: Graph Spatio-Temporal Networks for Condition Monitoring of Wind Turbine, IEEE Transactions on Sustainable Energy, 2024.

Kingma, D. P., Welling, M., et al.: Auto-encoding variational bayes, 2013.

Li, X., Cheng, C., and Peng, Z.: Unsupervised construction of health indicator for rotating machinery via multi-criterion feature selection and attentive variational autoencoder, Science China Technological Sciences, 67, 1524–1537, 2024.

Liu, J., Yang, G., Li, X., Hao, S., Guan, Y., and Li, Y.: A deep generative model based on CNN-CVAE for wind turbine condition monitoring, Measurement Science and Technology, 34, 035 902, 2022.

Liu, X., Teng, W., Wu, S., Wu, X., Liu, Y., and Ma, Z.: Sparse dictionary learning based adversarial variational auto-encoders for fault identification of wind turbines, Measurement, 183, 109 810, 2021.

Pandit, R. and Wang, J.: A comprehensive review on enhancing wind turbine applications with advanced SCADA data analytics and practical insights, IET Renewable Power Generation, 18, 722–742, 2024.

Peng, D., Desmet, W., and Gryllias, K.: Reconstruction-based deep unsupervised adaptive threshold support vector data description for wind turbine anomaly detection, Reliability Engineering & System Safety, 260, 110 995, 2025.

Ren, Z., Verma, A. S., Li, Y., Teuwen, J. J., and Jiang, Z.: Offshore wind turbine operations and maintenance: A state-of-the-art review, Renewable and Sustainable Energy Reviews, 144, 110 886, 2021.

Rezamand, M., Kordestani, M., Carriveau, R., Ting, D. S.-K., Orchard, M. E., and Saif, M.: Critical wind turbine components prognostics: A comprehensive review, IEEE Transactions on Instrumentation and Measurement, 69, 9306–9328, 2020.

Rolinek, M., Zietlow, D., and Martius, G.: Variational autoencoders pursue pca directions (by accident), in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 12 406–12 415, 2019.

Sandoval, D., Leturiondo, U., Vidal, Y., and Pozo, F.: Entropy indicators: An approach for low-speed bearing diagnosis, Sensors, 21, 849, 2021.

Su, X., Shan, Y., Li, C., Mi, Y., Fu, Y., and Dong, Z.: Spatial-temporal attention and GRU based interpretable condition monitoring of offshore wind turbine gearboxes, IET Renewable Power Generation, 16, 402–415, 2022.

Sun, S., Li, Q., Hu, W., Liang, Z., Wang, T., and Chu, F.: Wind turbine blade breakage detection based on environment-adapted contrastive learning, Renewable Energy, 219, 119 487, 2023.

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, 4381, 2022.

https://doi.org/10.5194/wes-2025-123
Preprint. Discussion started: 6 August 2025
© Author(s) 2025. CC BY 4.0 License.

Wang, A., Qian, Z., Pei, Y., and Jing, B.: A de-ambiguous condition monitoring scheme for wind turbines using least squares generative adversarial networks, Renewable Energy, 185, 267–279, 2022.

Wang, S., Vidal, Y., and Pozo, F.: An unsupervised approach to early fault detection and performance degradation assessment in bearings, Advanced Engineering Informatics, 68, 103 620, 2025a.

Wang, Z., Zhang, M., Chen, H., Li, J., Li, G., Zhao, J., Yao, L., Zhang, J., and Chu, F.: A generalized fault diagnosis framework for rotating machinery based on phase entropy, Reliability Engineering & System Safety, 256, 110 745, 2025b.

WeDoWind: The RTDT Research Affiliate Programme Space, https://www.wedowind.ch/spaces/the-rtdt-research-affiliate-programme-space.

Wu, P., Wang, Y., Zhang, X., Gao, J., Wang, L., and Liu, Y.: Wind Turbine Blade Breakage Monitoring with Mogrifier LSTM Autoencoder, IEEE Transactions on Instrumentation and Measurement, 2023.

Xu, H., Pang, G., Wang, Y., and Wang, Y.: Deep isolation forest for anomaly detection, IEEE Transactions on Knowledge and Data Engineering, 35, 12 591–12 604, 2023.

Xu, X., Huang, X., Bian, H., Wu, J., Liang, C., and Cong, F.: Total process of fault diagnosis for wind turbine gearbox, from the perspective of combination with feature extraction and machine learning: A review, Energy and AI, 15, 100 318, 2024.

Yan, X., Xu, Y., She, D., and Zhang, W.: Reliable fault diagnosis of bearings using an optimized stacked variational denoising auto-encoder, Entropy, 24, 36, 2021.

Yang, T., Jiang, L., Guo, Y., Han, Q., and Li, X.: LTFM-net framework: Advanced intelligent diagnostics and interpretability of insulated bearing faults in offshore wind turbines under complex operational conditions, Ocean Engineering, 309, 118 533, 2024.

Ying, W., Li, Y., Noman, K., Zheng, J., Wang, D., Feng, K., and Li, Z.: Stockwell transform spectral amplitude modulation method for rotating machinery fault diagnosis, Mechanical Systems and Signal Processing, 223, 111 884, 2025.

Yu, W., Kim, I. Y., and Mechefske, C.: An improved similarity-based prognostic algorithm for RUL estimation using an RNN autoencoder scheme, Reliability Engineering & System Safety, 199, 106 926, 2020.

Zhang, C., Hu, D., and Yang, T.: Anomaly detection and diagnosis for wind turbines using long short-term memory-based stacked denoising autoencoders and XGBoost, Reliability Engineering & System Safety, 222, 108 445, 2022.

Zhou, H., Huang, X., Wen, G., Lei, Z., Dong, S., Zhang, P., and Chen, X.: Construction of health indicators for condition monitoring of rotating machinery: A review of the research, Expert Systems with Applications, 203, 117 297, 2022.