Renewable Energy 203 (2023) 161–176

Contents lists available at ScienceDirect

Renewable Energy

journal homepage: www.elsevier.com/locate/renene

Fault detection of offshore wind turbine drivetrains in different environmental conditions through optimal selection of vibration measurements

Ali Dibaj *, Zhen Gao, Amir R. Nejad

Department of Marine Technology, Norwegian University of Science and Technology (NTNU), 7491, Trondheim, Norway

ARTICLE INFO

Keywords:
Offshore wind turbine
Drivetrain system
Fault detection
Optimal vibration measurements

ABSTRACT

In this study, a vibration-based fault detection method is proposed for offshore wind turbine drivetrain based on the optimal selection of the acceleration measurements. The main aim is to find the sensor positions mounted on the drivetrain that provides the most fault-related information. In fact, this study tries to optimize the vibration sensors suggested by ISO standards in terms of their position and number in order to get accurate fault detection results. The faults are considered in a set of bearings with a high probability of failure in a 5-MW reference drivetrain high-fidelity model installed on a spar-type floating wind turbine. Different simulated shaft acceleration measurements are examined under three environmental conditions. Correlation analysis is first performed on the measurements to see how the different faults and environmental conditions affect the correlation between the measurements. Then, a combined principal component analysis (PCA) and convolutional neural network (CNN) is employed as the fault detection method through the optimal vibration measurements. The prediction findings demonstrate that only two vibration sensors, one near the main shaft and another near the intermediate-speed shaft, can fully detect the considered faulty bearings. Also, it will be shown that the axial vibration data give more promising results than the radial ones which can be used in virtual digital twin models.

1. Introduction

The demand for wind energy as a renewable economic source is increasing worldwide. According to the Global Wind Energy Council (GWEC), global installed capacity has reached 837 GW by the end of 2021 [1]. In the meantime, offshore wind energy technology is rapidly developing due to access to higher-quality wind power, the possibility to install larger wind turbines (WTs), and the prevention of onshore wind energy operation problems such as sound and visual pollution [2]. Compared to onshore WTs, offshore WTs are more vulnerable to extreme loads due to the stochastic ocean conditions and irregular waves [3]. These harsh conditions increase the likelihood of unexpected breakdowns and severe damage. However, due to limited access to these structures, their Operation & Maintenance (O&M) costs are significantly higher than those of onshore WTs. According to statistics [4], O&M costs account for approximately 20%–25% of the total cost of an offshore WT. However, this number is around 10%–15% for onshore WTs. Hence, WT solutions for offshore development demand higher reliability, safety, availability, and serviceability than onshore designs in order to enhance the profitability and competitiveness of such projects and hence the attractiveness to potential investors [5,6]. Failures in the drivetrain components, according to experience, contribute significantly to the WT's downtime. As a result, it is crucial to use optimum maintenance strategies on offshore WTs, particularly drivetrain components [5,7].

The idea behind condition monitoring is that an incipient defect is detected by investigating changes in the system's responses. This identification is based on assessing changes in the data trend measured from the system. Common data acquisition systems for condition monitoring of WT drivetrain include vibration [8–10], temperature, and oil particle measurement [11,12]. Because the system defects induce rapid changes in vibration signals, vibration-based condition monitoring is one of the standard methods for fault detection of the drivetrain [13]. Some standards such as ISO 10816-21 [14] and ISO 16079-2 [15] recommend the vibration measuring positions for the drivetrain. These typical measurement locations should be close enough to parts with a high failure rate, such as bearings and gears, in order to respond quickly to fault-induced changes in the system. The proposed sensor placements, directions, and numbers for the drivetrain are in such a way that they

* Corresponding author.
E-mail addresses: ali.dibaj@ntnu.no (A. Dibaj), zhen.gao@ntnu.no (Z. Gao), amir.nejad@ntnu.no (A.R. Nejad).

https://doi.org/10.1016/j.renene.2022.12.049
Received 31 August 2022; Received in revised form 26 November 2022; Accepted 12 December 2022
Available online 14 December 2022

should fully define the vibration of the machine. However, using all or several of the vibration measurements suggested by these standards would not be cost-efficient because the condition monitoring system is typically not integrated into the wind turbine's control system and is supplied by a third party. Therefore, it is necessary to optimize the vibration sensors mounted on the drivetrain in terms of their position and number in order to provide rich fault-related information while employing the minimum number of sensors.

In order to get the optimized location and number of vibration sensors, the first step can be studying the relationship between the measurements. The common information between the multiple sensors is a function of the system's health, and it is subjected to change when a failure occurs. For example, many research studies have used the correlation degree between the vibration measurements as a fault indicator in fault detection of rotating machinery. The fault detection of rotor equipment based on multi-sensor correlation analysis and deep learning was done by Bai et al. [16]. For fault pattern recognition of a nuclear power plant, the correlation analysis was performed by Peng et al. [17] for dimension reduction of features. Then, these features were used as input to a deep belief network. Zhu et al. [18] proposed a damage indicator for structural damage detection based on the cross-correlation function. Also, a fault diagnosis method has been proposed by Xiong et al. [19] for rotating machinery based on fusing the dimensionless indices and Pearson correlation coefficient. These research studies show a clear connection between existing faults in the system and the correlation of sensor data.

The first part of this paper involves performing a correlation analysis on the vibration measurements obtained from the drivetrain in both radial and axial directions and under different environmental conditions. To contain the most fault-related information, these measurements are chosen from the parts closest to the likely failure locations. This study is being conducted to evaluate two probable effects on the correlation between vibration measurements: (1) the effect of failure occurrence in different elements of the drivetrain and (2) the effect of change in the environmental conditions.

The second part of this work concerns data-driven fault detection of the drivetrain using different acceleration measurements considering the effect of environmental conditions and measurement correlations. In recent years, different types of data-driven fault detection methods have been proposed with application to wind turbines. Zare and Ayati [20] proposed a multichannel CNN network for simultaneous fault diagnosis of 5 MW wind turbine benchmark model. They studied different types of simulated faults such as rotor imbalance and pitch actuator fault for this purpose. Li et al. [21] combined a convolutional autoencoder model with transfer learning strategy for fault diagnosis of wind turbine with small-scale SCADA data. A Takagi–Sugeno fuzzy model is used by Liu et al. [22] to address fault estimation and signal compensation for fault-tolerant control in a 4.8 MW wind turbine benchmark system. A multichannel fault diagnosis method for wind turbine driving system was proposed by Yan et al. [23] based on the multivariate singular spectrum decomposition and improved Kolmogorov complexity. Lei et al. [24] proposed a fault diagnosis framework for wind turbine based on an end-to-end LSTM model, without reliance to expert knowledge and feature selection through signal processing. In another work, a wind turbine fault detection method based on SCADA data and RNN model is suggested by Cui et al. [25]. However, the contribution of most of the works proposed recently is mainly related to the applied data-driven methods, and they have not been efficiently adapted to wind turbine applications, especially the drivetrain part.

In this study, for drivetrain fault detection using multi-point acceleration measurements, a combined principal component analysis (PCA) and convolutional neural network (CNN) approach is employed after determining how the measurements interact in different conditions. This deep learning-based fault detection method was recently used by the authors of this paper [26]. The same reference drivetrain model was employed for fault detection in that research. The method, however, was only validated at the rated wind speed. Furthermore, it was not examined which and how many vibration measurements can provide the highest fault detection accuracy. In addition, the first principal component was only used as input data to the CNN model. In the current study, different principal components are investigated in order to find the optimum input features for the CNN model. The employed method first uses PCA to combine information between the multiple vibration measurements to reduce the input size and the model's computational burden. The PCA-projected data is then fed into CNN for classification and fault-pattern recognition. Different vibration sensors, as suggested by ISO standards, in both radial and axial directions, are considered input data for the CNN model.

In summary, the main contributions of this study are listed as follows:

- A fused data-driven PCA-CNN method is proposed for the fault detection of a 5 MW floating offshore wind turbine drivetrain through optimal selection of vibration measurements.
- The position and number of vibration sensors are optimally selected for the bearing fault detection under different environmental conditions to provide reliable fault-related information using minimum required data.
- Since a multi-point vibration sensor design is used in this study, a correlation analysis is conducted on the measurements in the axial and radial directions. This study gives an overview of whether and which vibration data can be used in the proposed fault detection method based on the degree of their correlation.
- This fault detection method can be utilized either for a real-world system or its virtual digital twin, where the online virtual sensors can be installed in any position and direction of the model.

The 5-MW reference gearbox [27] model is used for this research. This reference gearbox model is installed on a spar-type floating wind turbine. The gearbox is modeled in the multibody simulation (MBS) environment in SIMPACK software. Since failures in the drivetrain typically initiate with bearings [28], bearing damages are considered in this study. The fault cases are considered in the main, high-speed shaft, and planet bearings. A de-coupled analysis approach has been pursued in this study. So that the forces and torques obtained from the global analysis (using SIMO-RIFLEX-AeroDyn simulation tools [29]) are used as input to the drivetrain MBS model at three different environmental conditions (ECs). Then, the internal dynamic responses of the drivetrain model, including the axial and radial acceleration measurements obtained from the main, low, intermediate, and high-speed shafts, are investigated as condition monitoring data.

The remainder of the paper is structured as follows: Section 2 describes the configuration of the wind turbine and drivetrain model used in this study, the de-coupled approach used to get loads applied on the drivetrain model, as well as the fault cases, and how to simulate faults in the MBS model. Section 3 discusses the methodology employed in this study, including correlation analysis, PCA, and CNN model descriptions. Section 4 contains the results and discussion. Finally, the conclusion is outlined in Section 5.

2. Numerical models

2.1. Wind turbine and drivetrain model

A 5-MW reference gearbox [27] installed on the floating OC3 Hywind spar structure [30,31] is used in this study. The spar floating structure is column-shaped and fastened to the seabed by three mooring lines. This WT is a 3-blade upwind turbine with a rated wind speed of 11.4 rpm. The wind turbine specification is provided in Table 2, while the overall characteristics of the spar floating platform are summarized in Table 1.

Table 1
Wind turbine specification [30].

Parameter	Value
Type	Upwind/3 blades
Cut-in, rated, and cut-out wind speed (m/s)	3,11.4 and 25
Rated rotor speed (rpm)	12.1
Hub height (m)	90.0
Rotor diameter (m)	126
Hub diameter (m)	3
Rotor mass (×1000 kg)	110
Nacelle mass (×1000 kg)	240
Hub mass (×1000 kg)	56.8
Wind turbine control system	Variable speed and pitch

Table 2
Spar floating platform characteristics [31].

Parameter	Value
Water depth (m)	320
Displacement (tonne)	8227
Hull mass (tonne)	7466
Draft (m)	120
Waterline diameter (m)	6.5

Table 3
5-MW reference gearbox specification [27].

Parameter	Value
Type	2 Planetary + 1 Parallel
1st stage ratio	1:3.947
2nd stage ratio	1:6.167
3rd stage ratio	1:3.958
Total ratio	1:96.354
Designed power (kW)	5000
Rated input shaft speed (rpm)	12.1
Rated generator shaft speed (rpm)	1165.9
Rated input shaft torque (kN m)	3946
Rated generator shaft torque (kN m)	40.953
Total dry mass (×1000 kg)	53
Service life (year)	20

The 5-MW reference gearbox used in this study has been developed by Nejad et al. [27] for offshore WTs utilizing commercial software, SIMPACK [32]. This gearbox resembles the most common design types used in WTs which includes three stages, two planetary, and one parallel stage gears. In order to restrict non-torque loads entering the gearbox, a 4-point support configuration with 2 main bearings is considered for this model. In this reference model, bearings are modeled using the SIMPACK force element with their stiffness values. Gears are also modeled using the SIMPACK gear pair element with consideration of the tooth geometry and modifications. Table 3 displays the general specifications for this gearbox.

2.2. De-coupled approach and environmental conditions

In this paper, to get the internal dynamic responses of the drivetrain model in SIMPACK, the results from the global analysis of the spar floating WT (applied forces and torques at the tower top) using the SIMO-RIFLEX-AeroDyn tools [29] are used as input to this model. Global and local analyses are conducted separately in this de-coupled approach due to the difference in the frequency ranges for the external and internal excitations of the drivetrain model. The frequency range for the external excitations – including wind and wave – is usually less than 2 Hz. External excitations also involve the dynamic loads on the main shaft caused by the vibration of the entire wind turbine system, such as tower bending or blade bending. On the other hand, internal excitations are usually within high-frequency ranges (gear mesh frequencies, for example). Therefore, simulations with high sampling frequencies (200 Hz) are needed to capture internal resonances. In contrast, simulations with small sampling frequencies (10 Hz) will also be able to capture possible responses to external excitations. The difference in these simulation frequencies in global and local analysis and the computational complexity for the drivetrain model require using a de-coupled analysis approach.

Table 4
Environmental conditions [33].

	EC1 (below-rated)	EC2 (rated)	EC3 (above-rated)
Wind speed U (m/s)	7.0	12.0	14.0
Turbulence intensity I (–)	0.19	0.15	0.14
Significant wave height H_s (m)	4.5	5.0	4.0
Spectral peak period T_p (s)	12.0	12.0	10.0

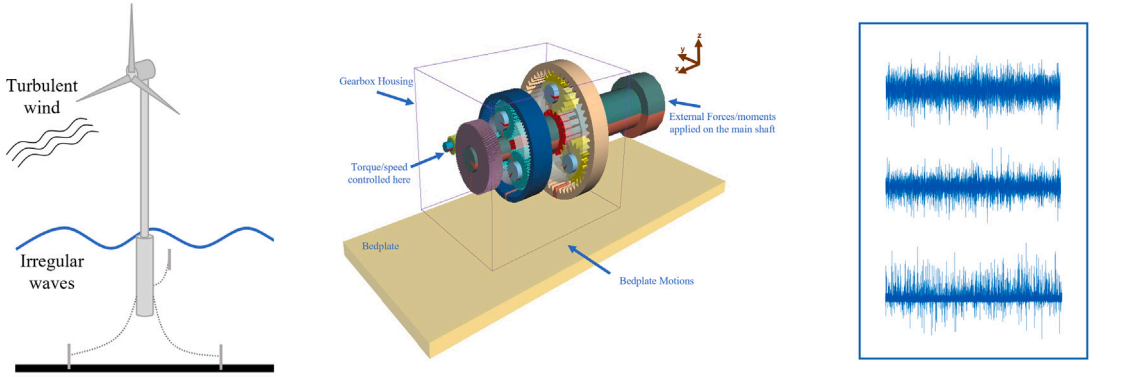
Numerical simulations for the global analysis of spar floating WT were conducted using aero-hydro-servo-elastic code by Nejad et al. [33]. Three environmental conditions (ECs) investigated in [33] are selected for this study as shown in Table 4 describing the characteristic of wind (hub-height mean speed U and turbulence intensity I) and wave (significant wave height H_s and peak period T_p). The reader is referred to Nejad et al. [33] for more information about global analysis and the parameters used in this analysis.

After global analysis, the acquired forces and moments are used as input into a detailed gearbox system modeled in the MBS environment, where simulations are performed at a higher sampling frequency (200 Hz). Different fault conditions (FCs) are considered for the drivetrain. For each of these FCs, a 3800-s simulation is conducted. The first 200 s are removed during the post-processing to prevent temporary start-up effects. From the acquired internal responses in the MBS model, axial and radial acceleration measurements obtained from the main ($A1A$, $A1R$), low ($A2A$, $A2R$), intermediate ($A3A$, $A3R$), and high-speed ($A4A$, $A4R$) shafts are considered as condition monitoring data for fault detection (see Fig. 2). Since the housing of bearings is not modeled in this drivetrain, the acceleration measurements from the shaft bodies (closest bodies to the bearing elements) in the MBS model are selected as the condition monitoring data.

Fig. 1 shows the overall procedure of the de-coupled analysis approach and the detailed model of the 5-MW reference gearbox. As seen in Fig. 1, the motions are applied on the bed plate and the external loads on the main shaft. The generator torque and speed are then controlled at the generator side [27].

2.3. Fault cases and fault modeling

As the core components of the rotating machinery, rolling bearings have an important role in the reliable operation of the equipment. On the other hand, due to the harsh and variable loads and stresses on the bearings of gearboxes in wind turbines, failures in these components are much earlier than the expected life of 20 years [34]. For that reason, fault cases (FCs) are considered in a set of bearings in this study. The selected bearings are based on the drivetrain bearings vulnerability map proposed by Nejad et al. [27]. Different approaches for calculating bearing stiffness have been developed in recent years, including finite element [35] and analytical methods [36]. As the bearing degrades, material loss occurs, as does a change in surface deformation and hardness of the bearing's internal components. As a result, the contact zone expands, and the bearing stiffness decreases. For example, new bearings typically have very high stiffness values (in the order of 10^8). This value, however, gradually decreases as the bearing degrades [37]. This bearing property enables a damage simulation approach in the MBS modeling. As a result, bearing damage can be modeled by varying the bearing stiffness in axial or radial directions. The increase in the bearing vibration level can represent the physical meaning of the stiffness decrease because a faulty bearing is expected to cause an increase in vibration amplitude. Fig. 2 and Table 5 show the fault cases considered in this study. The original stiffness values and the reduced values corresponding to degraded bearings in each load case are also given in Table 5.



Step 1: Global analysis

Step 2: External loads obtained by global analysis are applied on a detailed drivetrain model

Step 3: Signal processing applied on the condition monitoring data obtained from the internal dynamic of drivetrain model

Fig. 1. Drivetrain de-coupled analysis method.

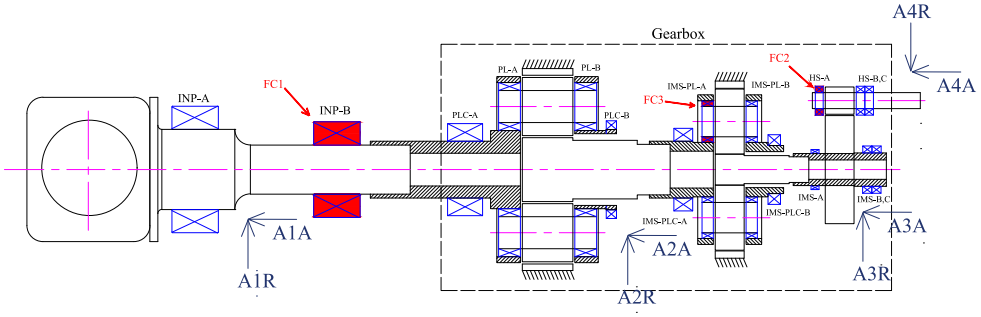


Fig. 2. Fault and measurement locations on drivetrain schematic layout.

Table 5
Drivetrain fault cases and corresponding stiffness values.

Fault case	Description	Original stiffness value (Healthy) (N/m)	Reduced stiffness value (N/m)
FC0	Reference case, healthy	–	–
FC1	Damage in axial direction of main bearing (INP-B)	4.06×10^8	4.06×10^7
FC2	Damage in radial direction of the high-speed shaft bearing (HS-A)	8.2×10^8	8.2×10^6
FC3	Damage in radial direction of the low-speed shaft bearing (IMS-PL-A)	6.12×10^7	6.12×10^5

3. Methodology

3.1. Correlation analysis

Correlation analysis is used in this paper to investigate the relationship of the responses obtained from the drivetrain system. This relationship is expected to change in the case of failure or different environmental conditions. The parameter describing the relationship in correlation analysis is known as the correlation coefficient. The data obtained from rotating machinery vibration sensors are continuous variables. The Pearson correlation coefficient is widely used in the literature for correlation analysis of continuous data. It has many advantages, such as high efficiency, accurate calculation, robust practicability, etc. This value is used to quantify linear relationships and also requires the data to conform to a normal distribution [16,17]. As a result, the Pearson correlation coefficient is used in this study to calculate the correlation between the acceleration measurements.

For the two variables (acceleration measurements in this study) $A1$ and $A2$, the Pearson correlation coefficient is defined as follows [38]:

$$\rho = \frac{E[(A1 - \mu_{A1})(A2 - \mu_{A2})]}{\sqrt{E[(A1 - \mu_{A1})^2]E[(A2 - \mu_{A2})^2]}} \tag{1}$$

where $E(.)$ is the expectation value. μ_{A1} and μ_{A2} are the mean values of $A1$ and $A2$, respectively.

3.2. Principal component analysis (PCA)

PCA is a linear statistical method based on multivariate analysis of the covariance matrix [39]. PCA extracts linear features from multivariate input data using dimension reduction. Input data is represented in another space called feature space by a set of orthogonal and independent components known as principal components, with the first principal component containing the highest possible variance of the input data. Similarly, the variance of the proceeding principal components decreases in descending order.

For a set of multivariate $n \times m$ input data $X = [A1, \ldots, A4]$, where n is the number of observations (length of acceleration measurement vectors = 2000) and m is the number of dimensions or variables (number of acceleration measurements), the PCA procedure is as follows:

1. Normalize the input data X:

$$Z_{ij} = (x_{ij} - \mu_j)/\sigma_j \tag{2}$$

where indices i and j are corresponding indices for the row and column vectors of input matrix X, respectively. x_{ij} also represents one element of matrix X. μ_j and σ_j are also column vectors' mean and standard deviation, respectively.

2. For Z, compute the $m \times m$ covariance matrix C.

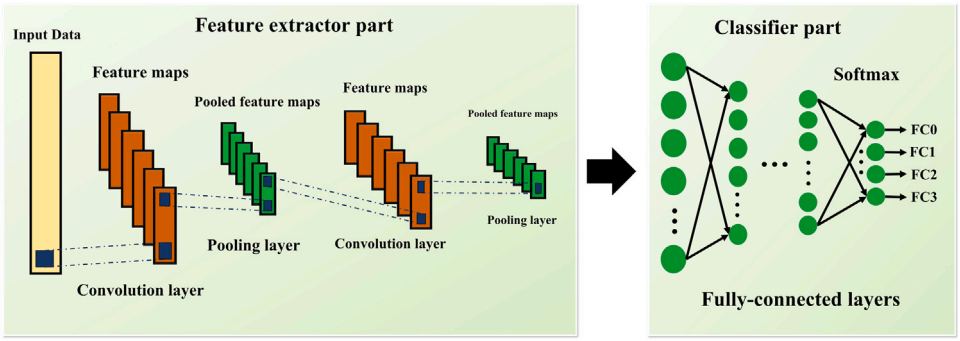


Fig. 3. CNN architecture.

3. Compute the eigenvalues λ and eigenvectors Q of the covariance matrix C:

$$Cq = \lambda q \longrightarrow \begin{cases} \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m \\ Q_{m\times m} = [q_1, q_2, \ldots, q_m] \end{cases} \quad (3)$$

where each column of Q represents one principal component coefficients (eigenvector) for the corresponding eigenvalue.

4. Project original space of input data X onto the principal component space:

$$P = Q^T X \quad (4)$$

Both P and X have the same dimension $n \times m$. P contains the principal component scores. Principal component scores are the representations of original data X in the principal component space.

3.3. Convolutional neural network (CNN)

CNN networks consist of two parts, as illustrated in Fig. 3: feature extractor and classifier. First, high-level features called feature maps are extracted from the input data using multiple convolution and pooling layer pairs in the feature extractor part. The classifier part, which consists of multiple fully connected layers, then assigns the extracted features to the corresponding correct output. Since, in this study, input data for CNN are one-dimensional vectors, which are the outputs from PCA, 1D CNNs are introduced in the following. For one-dimensional CNNs, It is assumed that the input of the convolution layer is $B \in^{W\times 1}$ where W is the length of one-dimensional input data B. The output of the convolution layer is calculated by Eq. (5) [40]:

$$M_j = \phi(B * W_j + b_j) \quad (5)$$

where $*$ denotes the convolution operator. M_j, W_j, and b_j are jth acquired feature map, weight matrix of jth convolution filter, and jth current layer bias, respectively. Also, ϕ is a nonlinear activation function that is generally the Rectified Linear Unit (ReLU), Hyperbolic Tangent (Tanh), or Sigmoid.

The pooling layer also reduces the dimensions of resulting feature maps to reduce the computational burden and to prevent overfitting. This layer reduces the size of feature maps by combining the neighbor values in the feature map and giving them a unique value. The pooling process is performed by various methods such as mean, maximum, and random pooling [41]. For example, in the max pooling, which is the most common pooling method, in a specified dimension ($S\times 1$) (pooling dimension) of the one-dimensional feature map M_j, the maximum value is selected as the pooled one:

$$K_j{}^i = \max_{PS\in S\times 1}(M_j^{PS}) \quad (6)$$

where M_j^{PS} is a part of feature map with dimensions of $PS = S \times 1$. Also, K_j^i is the ith element of the jth pooled feature map.

The output vectors of different pooling layers are gathered and flattened in a single vector to enter the classifier. The operation in this part is similar to that of a multilayer perceptron (MLP) network. The output layer typically uses the Softmax activation function for multi-class classification. This activation function returns the conditional probability for each class as output. Assuming that there are N classes to classify the data, the probability of output corresponding to the kth node or class ($O_k \in [0, 1]$) in the output layer with the Softmax function is calculated as follows:

$$O_k = P(y = k|x; W_k, b_k) = Softmax(W_k x + b_k) = \frac{exp(W_k x + b_k)}{\sum_{i=1}^{N} exp(W_i x + b_i)} \quad (7)$$

where $P(.)$ is the conditional probability function, x is the input of output layer and $\sum_{i=1}^{N} O_i = 1$. The output node corresponding to the maximum value of the obtained conditional probability is considered the input data class.

In order to minimize the error between the network output and the corresponding correct output (target), the categorical cross-entropy cost function is used as Eq. (8) [40]:

$$J = \sum_{i=1}^{N} -d_i ln(y_i) - (1 - d_i) ln(1 - y_i) \quad (8)$$

where y_i and d_i are the network and correct outputs, respectively. Adam's optimization algorithm is also used to minimize the cost function.

3.4. Construction of fault detection model

This study performs the drivetrain fault detection method in a pattern recognition framework. The fault detection method in this framework consists of three steps: data acquisition and preparation, extraction of fault-related features, and classification [42]. For the first step, the numerical condition monitoring data are acquired from the drivetrain simulated model in SIMPACK in acceleration measurements in three different environmental conditions, as described in Section 2.2. Then, these measurements are segmented into shorter-length signals to create enough data for training and testing the machine-learning model. In the next step, PCA is applied to transform the feature space of the original measurements into the principal components' feature space. As explained in Section 3.2, in the principal components feature space, variables are sorted based on their variability in the principal components feature space. One could use all the measurements as monitoring data or the first few principal components. The latter may give the same prediction accuracy because the first few principal components can present significant variability of the original feature space. The computational burden is also significantly reduced using the first few principal components. The classification step also includes feeding the monitoring data, either raw measurements or PCA-projected responses, into the CNN model to classify different fault conditions. A schematic view of the fault detection model is depicted in Fig. 4.

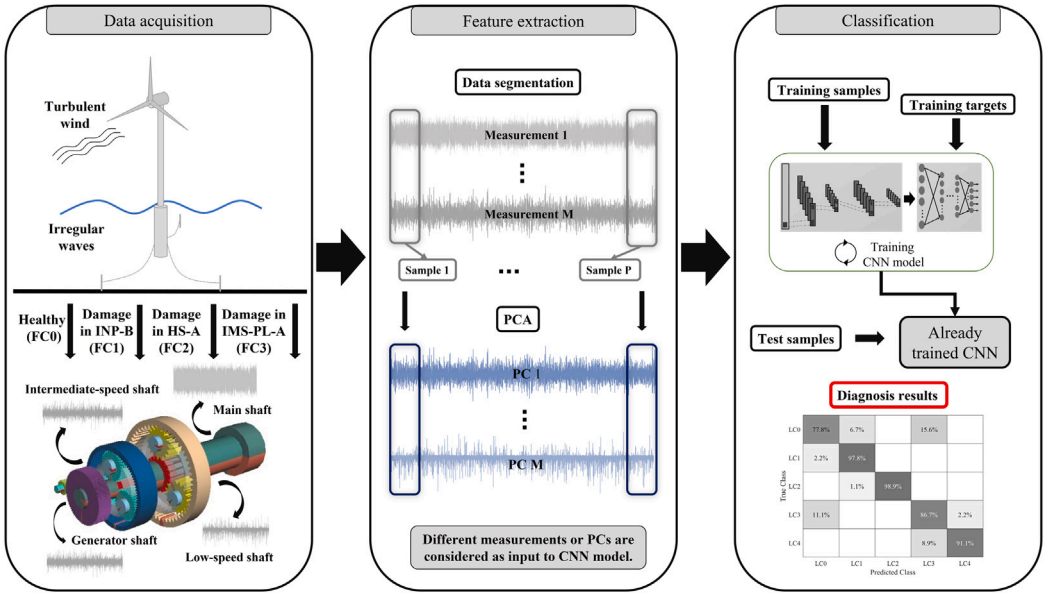


Fig. 4. Schematic description of the proposed method for the fault detection of the drivetrain.

4. Results and discussion

4.1. Data preparation

As explained in the previous section, to have enough training and test data for the data-driven method and reliable correlation analysis of the measurements, it is necessary to segment 1-h measurements obtained in different fault conditions. As an example of 1-h simulations, for EC1 and FC1, four 1-h axial acceleration measurements obtained from considered locations are shown in Fig. 5. These 1-h measurements are segmented into shorter-length signals (10 s). For each fault condition, 360 segmented signals are considered in each axial and radial direction (1440 signals in total for each direction and environmental condition), each with a length of 2000 sample points. It should be noted that there will be two separate sets of simulations, one for training and one for test dataset for the CNN model in order to decrease the likelihood of model overfitting. However, each 1-h simulation is itself made up of six 10-min individually modeled wind distributions. It should also be noted that in this study, each environmental condition and also each direction are considered separately in the analysis (either correlation analysis or fault detection), and the number of signals mentioned is for one of the ECs and directions and training or test datasets.

The segmented measurements should contain the required information from the system. On the other hand, they should be statistically different for reliable predictions and prevention of overfitting the machine learning model. To show the statistical variation between the segmented signals for each acceleration measurement and under different fault conditions, the one-way ANOVA statistical test is employed with 95% confidence level ($\alpha = 0.05$). The ANOVA results are only shown for EC2 (rated wind speed) because of the limited space in the paper. The results in terms of *p*-values for axial and radial directions are given in Tables 6 and 7, respectively. Statistically significant difference is clearly seen from the *p*-value results, especially for the axial direction.

4.2. Correlation analysis results

In this section, correlation analysis is conducted between the acceleration measurements to investigate their correlation patterns and how they change in the case of different faults and environmental conditions. The correlation analysis results are presented between the acceleration responses acquired from the main ($A1$), low-speed ($A2$), intermediate-speed ($A3$), and generator ($A4$) shafts in both axial and radial directions.

Table 6
One-way ANOVA statistical test between the segmented measurements in axial direction for EC2 and each fault condition.

	p-value			
	FC0	FC1	FC2	FC3
$A1A$	0	0	0	0
$A2A$	0	0	0	0
$A3A$	0	0	0	0
$A4A$	0.0026	0.0006	0.0001	0.0004

Table 7
One-way ANOVA statistical test between the segmented measurements in radial direction for EC2 and each fault condition.

	p-value			
	FC0	FC1	FC2	FC3
$A1R$	0.0184	0.032	0.0071	0.0037
$A2R$	1.86×10^{-8}	1.29×10^{-5}	0.002	0.003
$A3R$	0.0014	0.0026	0.002	0.0022
$A4R$	0.0118	1.34×10^{-5}	0.008	0.0032

The averaged values of absolute correlation coefficients between the segmented signals of axial and radial acceleration measurements obtained in different fault conditions are given in Fig. 6(a), (b), and (c) for EC1, EC2, and EC3, respectively. As can be seen from Fig. 6(a), (b), and (c), the correlation coefficient values in the axial direction are higher than the values obtained in the radial direction. The radial responses are completely uncorrelated. The lack of correlation in the radial direction can be attributed to the existing internal excitations – mesh frequencies, for instance – in the radial responses. These excitations differ from one stage to the next. As a result, the frequency content of the radial responses can also vary from stage to stage, resulting in uncorrelated radial responses. While in the axial direction, a significant correlation is seen. This high correlation can be due to two reasons: (1) Wind and wave-induced excitations applied to the drivetrain in the axial direction. External excitations, low-frequency components with large amplitudes, appear in all four axial vibration responses, resulting in high correlations, especially between $A1A$, $A2A$, and $A3A$. (2) The series positioning of the shafts in the drivetrain system in the axial

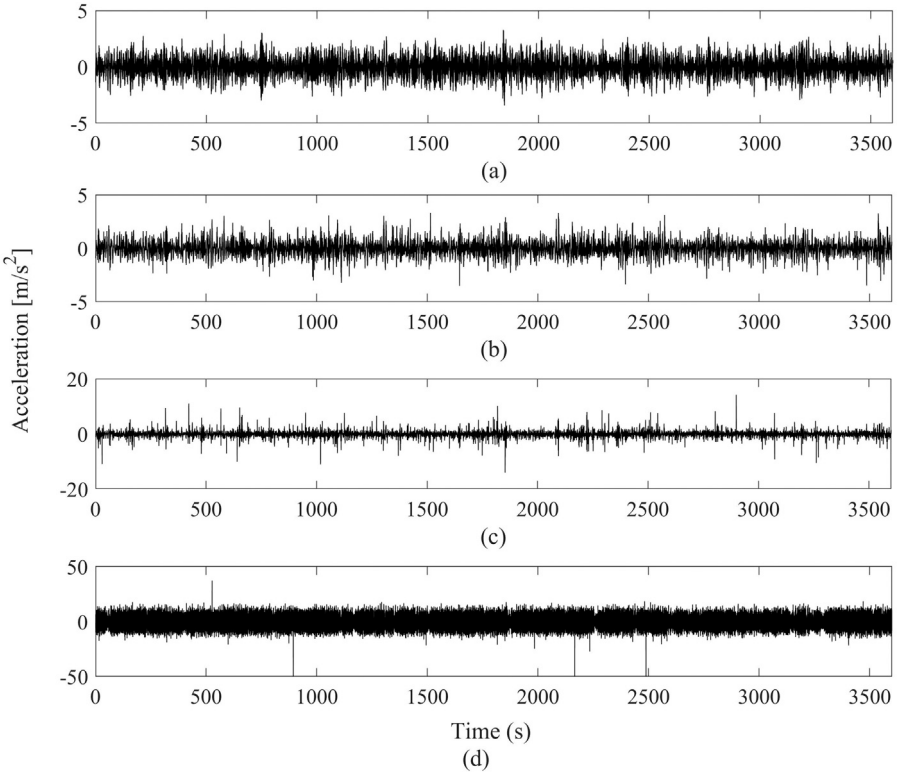


Fig. 5. 1-h axial acceleration measurements under EC1 and FC1 obtained from 5-MW gearbox model (a) Main (b) Low-speed (c) Intermediate-speed, and (d) High-speed shaft measurements.

direction causes a high correlation between their responses. As can be seen from the results in Fig. 6, the $A4A$ response from the generator shaft (non-series shaft with other ones) is almost uncorrelated with other responses. Also, as the wind speed increases, correlation values generally decrease. As in EC1 (Fig. 6(a)), measurements are more correlated compared to EC2 and EC3. There is less background noise in the acceleration measurements at lower wind speeds because of the low rotational speed of the main shaft. Therefore, the correlation degree of measurements is higher in lower wind speeds than in upper ones.

Concerning the fault conditions' effect on correlation coefficient values between the vibration responses, it is expected that when a fault occurs in each component, the correlations of the response close to that component change with the other responses [43,44]. Investigating the results of Fig. 6(a), (b), and (c), it is observed that the highest correlation values in all three wind speeds are obtained for the healthy condition (FC0). Also, it is clearly seen that with the occurrence of the faults, the correlation values generally decrease between the measurement close to the fault location and others. For FC2 (fault in the high-speed shaft bearing), this state is valid for all three wind speeds because of the location of the fault on the high-speed side and its great impact on the high-speed shaft measurement. As the fault locations move to the low-speed side of the drivetrain, the change in the correlation values decreases. For FC3 (fault in the low-speed shaft bearing), correlation values in the highly correlated measurements remain unchanged at EC1. However, for EC2 and EC3, this change is considerable. For FC1 (fault in the main bearing), no change in the correlation values is seen either for EC1 or EC3 because of the very low speed of the faulty bearing. But, at EC2, where the wind turbine operates at rated wind speed, there is a noticeable change in the correlation values of highly correlated measurements for FC1. This is the wind speed where the pitch control system becomes activated and often switches between modes due to the turbulent incoming wind. At this speed, the peak thrust force is reached based on the study done by Nejad et al. [45]. This leads to higher loads with a higher number of cycles and thus more significant fatigue damage. Therefore, the highest equivalent load occurs for INP-B bearing around this speed. Consequently, the vibration response related to this bearing would have more background noise and high-amplitude frequency components.

Table 8
The summary details of different inputs to CNN model.

CNN input	Input size
Individual raw acceleration measurements (see Fig. 7)	[2000 × 1]
Combined raw acceleration measurements (see Figs. 7 and 11)	[2000 × 2] or [2000 × 4]
Individual principal components (see Fig. 9)	[2000 × 1]
Combined principal components (see Fig. 9)	[2000 × 2] or [2000 × 4]

4.3. Fault detection of drivetrain system

The main objective of this study, as stated in the Introduction Section, is to find the optimized location and number of vibration sensors for the fault detection of the drivetrain system according to the suggested sensors by ISO standards. The PCA approach is used to first assess the selected measurements in order to keep the most information in the measurements while reducing the data dimension (number of measurements). In this regard, it will be investigated which principal components best represent the original data that can better predict the drivetrain health condition. On the other hand, based on the correlation analysis discussed in the previous section, the following will demonstrate which measurements may be ignored and which combination of measurements can yield the best result in fault detection. This research will determine the optimal number and position of measurement sensors. In summary, different data will be considered input data to the CNN model as described in Table 8 to compare their prediction outcomes.

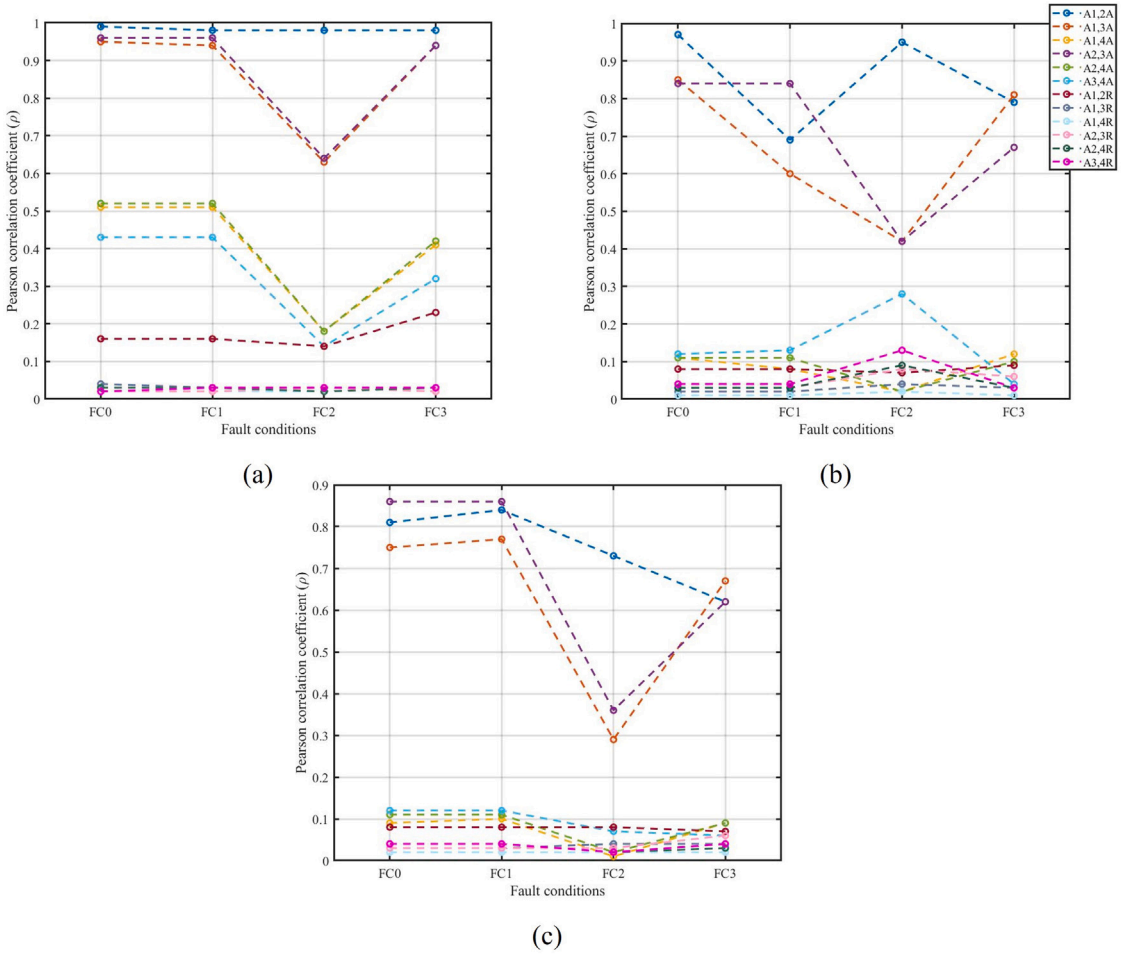


Fig. 6. Correlation values between the measurements in different fault and environmental conditions (a) EC1 (below-rated), (b) EC2 (rated), and (c) EC3 (above-rated).

As explained earlier, the deep 1D-CNN model is employed in this study for the fault classification task. CNN model detailed architecture is shown in Table 9. Various factors influence the process of determining and fine-tuning hyper-parameters, such as the number and size of convolution layers. For example, it is mainly determined by the input data size and the model performance. First, the training is started with a set of initial hyper-parameter values. The hyper-parameters are fine-tuned based on the obtained performance after each training-validation-test process until a convincing performance is reached. Furthermore, the arrangement of the convolution layers in the feature extractor part of the CNN model significantly affects the extracted features from the input data. A few convolution layers with larger sizes in the first layers can extract low-level features from input data. On the other hand, high-level features are formed as the size of convolution layers decreases, and their number increases in the final layers. This architecture significantly improves the computational burden while keeping good output accuracy. It should also be noted that the CNN model used in this study employs the stride strategy (stride length = 2) in convolution layers rather than adding extra pooling layers for downsampling. In this case, the convolution layer performs the convolution and downsampling simultaneously, resulting in a computationally efficient process.

In feeding this model, 80% of input data is used as training data, and the rest (20%) will be used as test data. This means that for each environmental condition, there will be 1440 train and test samples (segmented signals) in total, and 1152 samples will be considered for training (288 samples for each class), and 288 samples (72 samples for each class) will be used as test data. However, as mentioned before, training and test datasets come from different simulations. Also, 10% of training data will be used to validate the accuracy of CNN model at the end of each epoch (validation data). The length of input data to the CNN model will be 2000 in all cases, the same as the length of segmented signals. The output number of the CNN model would also be the same as the number of four classes (fault conditions). The CNN model is built and trained in Python using the keras library with the backend of Google's Tensorflow (Intel Core i7-10610U @ 1.80–2.30 GHz processor, 32 GB of RAM, and the Windows 10 operating system).

4.3.1. Axial and radial raw measurements as input data to CNN

In this section, the axial and radial raw measurements are directly fed into the CNN model for fault classification (without using PCA) to compare fault detection accuracy for the measurements of both directions. Fig. 7(a)–(f) show the classification results for both axial and radial measurements for three different wind speeds. According to the results, axial measurements provide better classification accuracy than radial ones. All classes can be classified with almost the highest accuracy in all ECs when all measurements are used. However, for radial measurements, the FC0 (healthy) and FC1 (main bearing fault) classes are misclassified in most cases. This poor identification might be related to how bearing faults are modeled. The fault in the main shaft bearing (FC1) is modeled axially. While the other two (FC2 and FC3) are modeled radially. Radial measurements can accurately detect radially modeled faults but cannot detect axial faults. Axial measurements, on the other hand, can detect both radial and axial faults. But the reason also comes back to the characteristic of acceleration

Table 9
Details of the 1D CNN architecture.

	Convolution kernels length/stride	Number of convolution kernels	Output vector length	Activation function
Input layer	–	–	2000 × 1	–
Convolution layer 1	(326 × 1)/2	8	1000 × 1	ReLU
Convolution layer 2	(163 × 1)/2	16	500 × 1	ReLU
Convolution layer 3	(81 × 1)/2	32	250 × 1	ReLU
Convolution layer 4	(27 × 1)/2	64	125 × 1	ReLU
Convolution layer 5	(9 × 1)/2	128	63 × 1	ReLU
Convolution layer 6	(3 × 1)/2	256	32 × 1	ReLU
Flatten	–	–	8192 × 1	–
Fully-connected layer 1	–	–	100 × 1	ReLU
Fully-connected layer 2	–	–	50 × 1	ReLU
Fully-connected layer 3	–	–	20 × 1	ReLU
Output layer	–	–	4 × 1	Softmax

measurements. They are more sensitive to high-impact rates with high-frequency components. According to the results, the challenge is mainly about fault detection of the main bearing (FC1). Test samples from this class are incorrectly classified as healthy class samples, and vice versa. This bearing is located on the main shaft (low speed). Low-speed side failures contain less impact energy. Because of the low-energy impact, there is a high noise ratio to an abnormal bearing signal [46]. Thus detecting the fault in this bearing using acceleration data, particularly radial, is difficult. Radial acceleration data mainly capture high-frequency internal dynamics of the drivetrain; in this dynamic, low-energy impacts are masked by other components. However, it is seen that among the radial measurements, A1R outperforms the others in fault detection of FC1. Since this measurement is located on the low-speed side of the drivetrain, it better captures low-speed fault-related information.

Based on these findings, it may be more appropriate to employ axial measurements to detect different fault locations in this 5 MW drivetrain model. However, the main question arises in this regard: mounting the sensors in axial directions may be challenging in the physical design of the drivetrain. This is why most CM systems recommend using radial sensors to collect vibration data. This problem can be handled by utilizing virtual digital twin models, where measurement sensors can be installed in any position of the virtual drivetrain. Therefore this recommendation can be considered for future studies on developing digital twin models.

In order to prove the effectiveness of the CNN model used in this study, fault detection results are compared with the results obtained using a multi-class support vector machine (SVM) classifier and multi-layer perceptron neural network (MLP). The models are compared only for the cases of all four axial measurements as input (results for the CNN model are already shown in Fig. 7(a),(c), and (e)). The compared models are shallow machine learning networks that need hand-crafted extracted features from raw data to train the network. In this regard, two features are considered as input to these models: (1) Four RMS values obtained from four acceleration measurements, (2) Six correlation values (ρ) between the four acceleration measurements. The multi-class SVM classifier consists of $K(K-1)/2$ binary SVMs (K is the number of classes), which use a one-versus-one coding design for classification. It should be noted that, for binary SVMs, Gaussian (RBF) kernel function is used, with the box constraint value set to one. Also, for MLP, three hidden layers with Sigmoid activation function and 100, 50, and 10 neurons are used, respectively. The output layer uses the SoftMax activation function with four neurons same as the number of classes or fault conditions. The division of segmented signals as training, validation, and test data is the same as for the CNN model. The classification results are shown as the detection accuracy for each class as well as the overall accuracy in Table 10 for different environmental conditions (ECs). As can be seen from Table 10, the CNN model outperforms the other two models using the raw axial measurements as input without using any hand-crafted feature.

4.3.2. Data fusion through PCA

In this section, PCA is applied to raw measurements before feeding into the CNN. As explained earlier, PCA is applied to reduce the dimension of the raw measurements while keeping the most informative data. PCA does this task by finding the orthogonal and independent directions (eigenvectors) of the original data space. Then these directions are sorted based on the data variance along each direction.

Since four acceleration measurements are acquired from the drivetrain working under each fault condition, the space of the original data would be four-dimensional. Hence, there will be four independent principal components (PC). Among the principal components obtained, it should be determined which holds the most information and should be kept and which should be discarded. In this regard, correlation analysis is performed between the principal components and the original measurements. The degree of correlation indicates how much information each component shares from each acceleration measurement. The correlation values are given in tables shown in Fig. 8(a)–(f). As shown, in the axial direction and for all ECs, the first and second PCs are the most correlated. The first PC mainly correlates with the A4A measurement (high-speed shaft acceleration data). The A4A measurement has a significantly greater variance (see Fig. 5(d)) than the others, especially at higher wind speeds. Furthermore, as seen in Fig. 6, this measurement is non-correlated with the others, particularly at higher wind speeds. As a result, the first PC mainly concentrates on information gathered from the A4A. The second PC is also correlated with the remaining three measurements. This means that the correlated information with these three measurements is embedded in the second PC. The obtained result was expected. Because according to the correlation analysis between the acceleration measurements given in Fig. 6, the three measurements, A1A, A2A, and A3A, had a high degree of correlation in all conditions. This high correlation led to the concentration of the correlated information between these three measurements in the second PC. Therefore, The first two PCs will be used as input to the CNN model.

In the case of radial measurements, Fig. 8(b), (d), and (f) show that each independent PC is focused on one independent measurement (sorted based on the standard deviation of radial measurements), such that each component is correlated with just one of the measurements. These results were also expected since, based on the correlation analysis done in the previous section, it was seen that the radial measurements are uncorrelated. Therefore, each uncorrelated measurement is distributed along one of the PCs. These results indicate that all principal components should be used as input to the CNN model when using radial data because each PC represents only one of the radial measurements due to the lack of correlation between them (see Fig. 6). As a result, using the PCA approach on radial data is not optimum since it does not reduce the data dimension, and as it will be shown, there will be no effect on the improvement of fault detection results.

In order to assess the effect of PCA on fault detection results, the correlated PCs with raw measurements are fed into the CNN model under all fault and environmental conditions for both axial and radial data. The correlated PCs (see Fig. 8) are only chosen because, as discussed, they contain the most information obtained from the raw

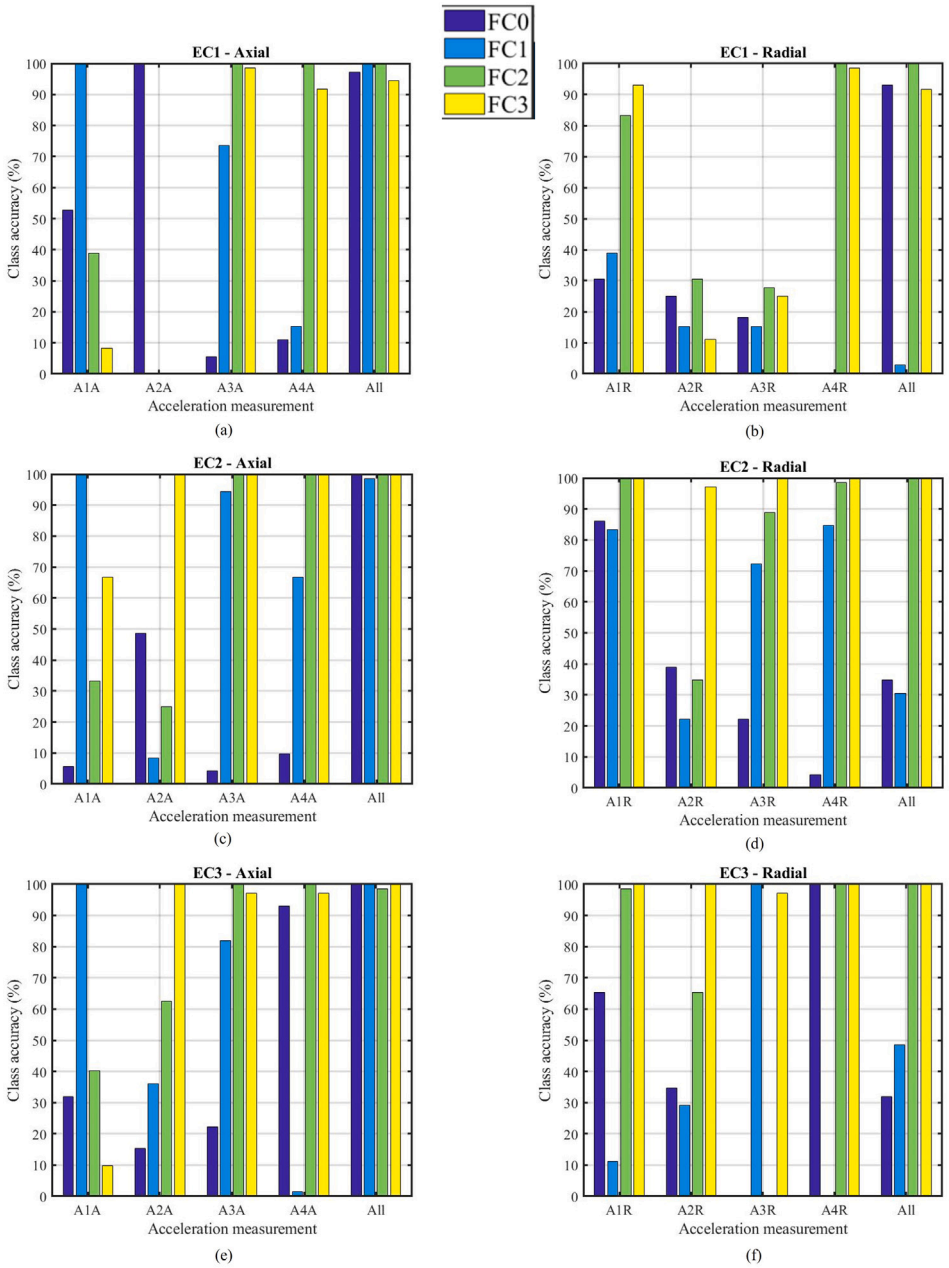


Fig. 7. CNN fault classification results with different raw acceleration inputs (a) EC1 - axial, (b) EC1 - radial (c) EC2 - axial, (d) EC2 - radial, (e) EC3 - axial, and (f) EC3 - radial.

measurements depending on the degree of their correlation with each of them. Therefore, the first two and all four PCs are selected for the axial and radial data, respectively. CNN prediction results are compared in two ways: (1) The PCs are used individually as input. (2) The combination of individual PCs is used as input. Fault classification results are shown in Fig. 9(a)–(f). The findings are as follows:

Axial: In all three wind speeds, the second PC obtains the highest prediction accuracy (100% accuracy is obtained in EC2 and EC3 for all fault conditions). Fig. 8 shows that PC2 has the highest correlation with the three measurements, A1A, A2A, and A3A. This correlation demonstrates that these three acceleration measurements contain the most system-related information under all fault conditions. The first PC, which has the highest correlation with A4A (particularly in EC2 and EC3), gives similar classification results to those obtained from A4A (see Fig. 7). It is noteworthy that the trained CNN with the combined first two PCs has a poorer prediction accuracy than the trained CNN with the individual second PC in EC1 and EC3 (prediction results are inclined to PC1 results). This can be due to the fact that when the first two PCs are combined, PC1 is the dominating component due to the more distributed feature space (higher variance). Also, it should be noted that a portion of information of the three correlated measurements is distributed in the next PCs, PC3 and PC4, based on the degree of their correlation with these measurements (see Fig. 8(a), (c), and (e)). Therefore, the percentage of shared information by these measurements in the combined first two PCs is lower than when the raw measurements are used. As a result, CNN will capture the most information from the first PC. Subsequently, the results will be quite identical to those of CNN trained with PC1 alone. These findings also highlight which measurements are suitable for condition monitoring data and which may be discarded without reducing the fault detection accuracy, as will be addressed further below.

Radial: Because of the lack of correlation between the radial measurements, each PC has just the information from one of the radial measurements, as shown in Figs. 6 and 8. In other words, the raw

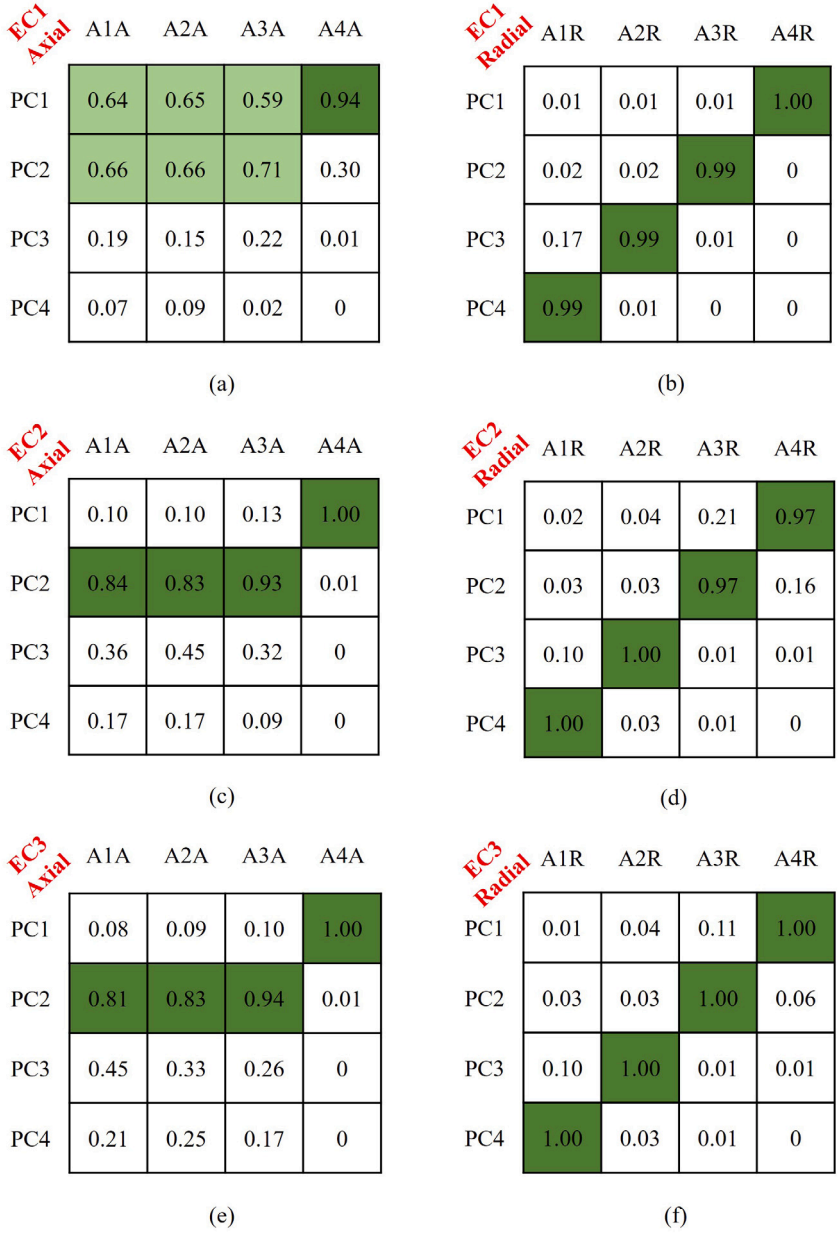


Fig. 8. Pearson correlation values between the principal components and raw acceleration measurements (a) EC1 - axial, (b) EC1 - radial (c) EC2 - axial, (d) EC2 - radial, (e) EC3 - axial, and (f) EC3 - radial.

radial measurements and principal components have a one-to-one correspondence. PC1, for example, has the highest correlation with the A4R sensor and so has the most information from this sensor. Because of this one-to-one relationship, the fault detection results are almost identical to the case in which the raw data was used (see the radial results in Figs. 7 and 9). No improvement has been achieved in data dimension reduction and classification accuracy. However, the principal components give different classification accuracies depending on their correlation with the measurements. For example, CNN trained with PC1 and PC4 (the principal components corresponding to A4R and A1R, respectively) outputs better classification results than the other two principal components. These are the radial accelerations in the input and output of the gearbox model. In general, it can be concluded that employing the PCA method and data fusion does not improve fault diagnostic accuracy in the case of radial data.

Finally, in this part, it is studied whether PCA and data dimension reduction may improve the convergence and training time of the CNN model. Actually, one of the primary motivations of this work in employing PCA is to check if PCA can reduce the computational burden and time required by CNN. Model convergence and training time are two important aspects to address in deep learning-based fault detection techniques. Deep learning models may require a long training time due to the multiple layers and the requirement for big data in real-world applications. In this context, preprocessing methods on raw data before feeding to the neural network can minimize the training time by reducing input data size. CNN training time is compared in two ways in this regard: (1) When PC2 is used as the input. (2) When all the axial measurements are used as input. Table 11 shows the average training times of CNN using these two inputs for the same number of training epochs. CNN takes somewhat less time to train with PC2 than it does with all measurements (22% reduction). However, because the input data size has decreased by a fourth, this reduction is less than predicted. Also, in order to show the convergence speed of the CNN model for both inputs, the loss and accuracy values obtained during the training of the model for EC4 are shown in Fig. 10. As can be observed,

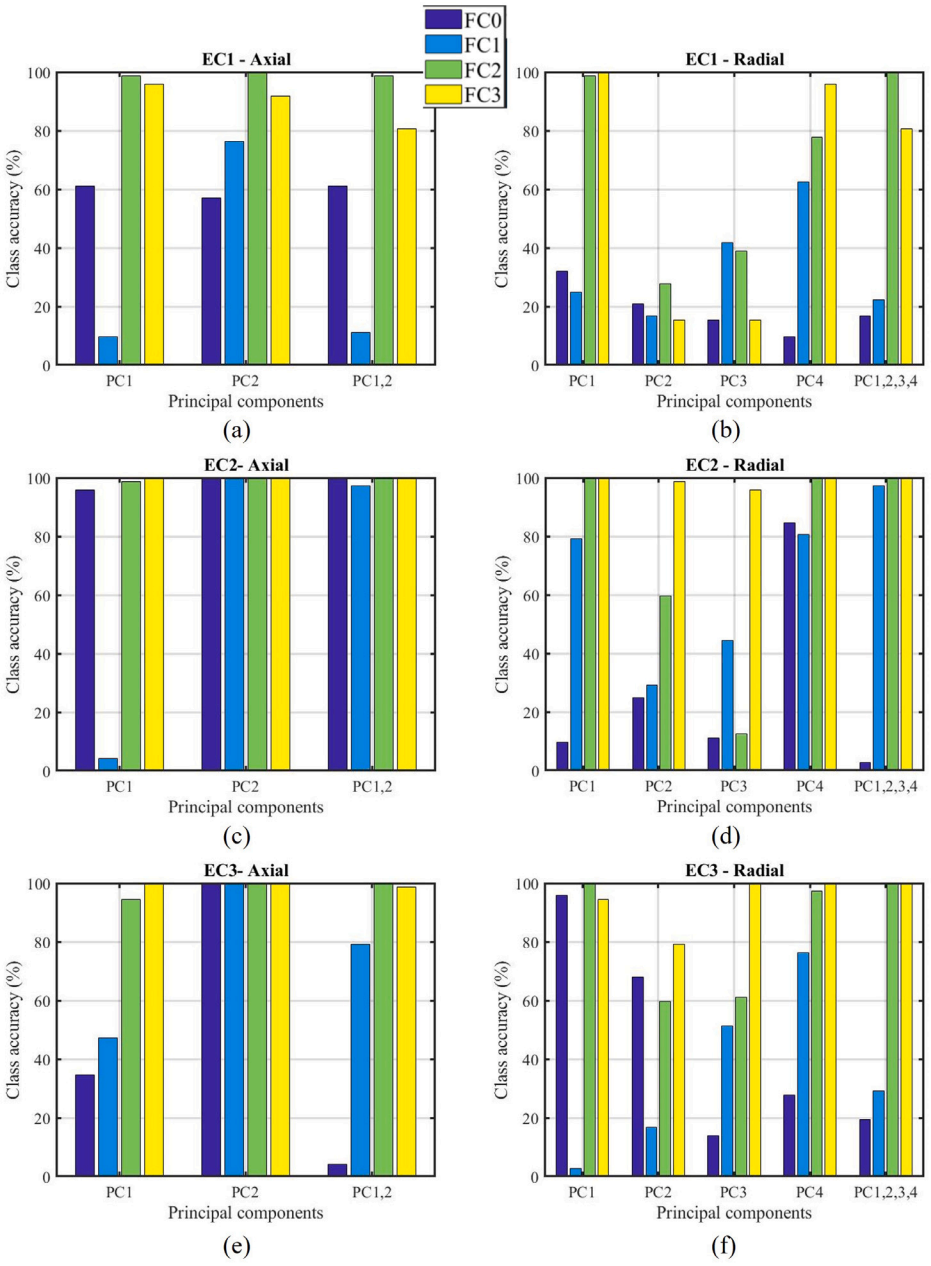


Fig. 9. CNN fault classification results with different PCs as input (a) EC1 - axial, (b) EC1 - radial (c) EC2 - axial, (d) EC2 - radial, (e) EC3 - axial, and (f) EC3 - radial.

the CNN model converges faster when PC2 is used as input than when all measurements are used. Because as previously stated, redundant information is removed, and the input data size is reduced through the PCA by keeping the most informative principal component. As a result, the CNN model converges in lower epochs. It should also be noted that using the PCA method and only the second PC, the same prediction accuracy is achieved as if all of the original axial measurements were utilized as the CNN input.

4.3.3. Position and number optimization of the axial measurements

Based on the results obtained in the previous section, it was shown that the second principal component obtained from the axial data provides the highest classification accuracy for the fault detection of the 5 MW drivetrain model with the considered bearing faults. PC2 is highly correlated with A1A, A2A, and A3A, implying that the majority of the information in this PC comes from these three correlated measurements. Based on this finding, it is concluded that the non-correlated A4A measurement is not required and that the correlated measurements are sufficient to achieve the best prediction accuracy for the considered fault conditions. One reason why A4A is not suitable for fault detection of the main bearing is that the first and second-stage bearings mainly capture the extra loads caused by the failure of this bearing. Therefore, this failure does not have much effect on the vibration of the high-speed shaft side (see also the detection results of A4A in Fig. 7).

It should now be investigated which combination of A1A, A2A, and A3A measurements gives the highest prediction accuracy. In this regard, the CNN model is fed a two-by-two combination of the three measurements A1A, A2A, and A3A. Classification results are shown in Fig. 11(a)–(c) for EC1, EC2, and EC3, respectively. As can be seen, the combination of A1A and A3A yields the maximum accuracy for all three ECs. This configuration is a two-sided sensor design, with one axial measurement from the main shaft side and one from the intermediate-speed side of the drivetrain model. The fused information from these two measurements contains the most fault-related characteristics and provides maximum fault prediction accuracy. As observed

Table 10
Fault classification results for three different machine learning models with all four axial measurements as input.

Environmental condition	ML model	Input feature	FC0 (%)	FC1 (%)	FC2 (%)	FC3 (%)	Overall (%)
EC1	SVM	RMS	15.3	2.8	2.8	73.7	23.7
		ρ	47.3	51.4	98.7	77.8	68.8
	MLP	RMS	36.9	19.2	33.4	63.1	38.1
		ρ	58.3	73.6	100	57.9	72.5
	CNN	–	97.22	100	100	94.44	98
EC2	SVM	RMS	83.4	94.5	100	95.9	93.4
		ρ	97.3	95.9	100	100	98.3
	MLP	RMS	100	94	100	100	98.5
		ρ	90.7	95.7	98.6	98.7	95.9
	CNN	–	100	98.61	100	100	99.7
EC3	SVM	RMS	37.5	27.8	100	87.5	63.2
		ρ	43.1	38.9	100	95.9	69.5
	MLP	RMS	70.7	48.8	100	86.7	76.6
		ρ	58.7	48.7	100	98.7	76.5
	CNN	–	100	100	98.61	100	99.7

Table 11
The training time of CNN model using different two inputs for 70 epochs.

CNN input	Training time (s)
Second principal component	415.96
All four axial measurements	537.02

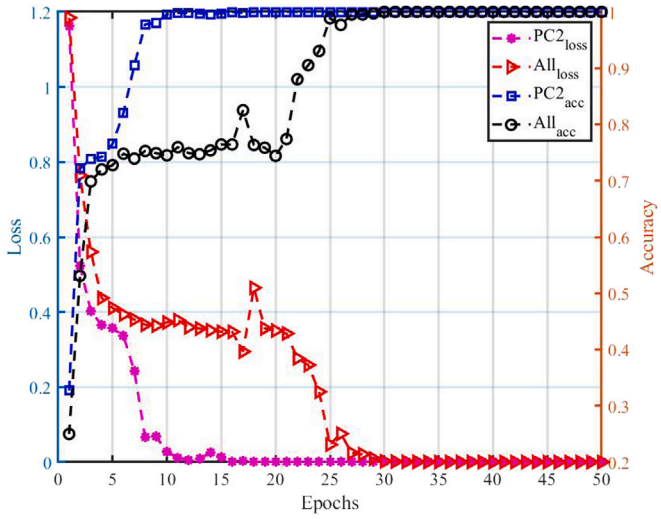


Fig. 10. Loss and accuracy values obtained during the training of CNN for two different inputs (a) Second principal component, and (b) All four axial measurements.

from Fig. 7(a), (c), and (e), A3A measurement provides the best classification accuracies between the individual axial measurements because of its quite central placement in the drivetrain system. The additional loads caused by all faulty bearings affect the vibration level of this measurement. However, the effect of FC1 is a bit less than the other two, as the detection accuracy for this fault condition is also less (Subsequently, the prediction accuracy of FC0 also decreases due to the misclassification of samples of two classes with each other). On the other hand, A1A could also detect the FC1 (main bearing fault) with 100% accuracy, while the other measurements could not. As a result, the combination of these two measurements, as well as their cross-correlation properties, enable a proper and optimized configuration for condition monitoring data acquisition. Therefore, based on this study, the pair of A1A and A3A measurements is recommended as the best sensor configuration among the suggested vibration measuring positions by ISO 10816–21 and ISO 16079–2 for the bearing fault detection of the 5 MW offshore wind turbine drivetrain model.

As a comparison study and to further prove the robustness of the CNN model, the classification results in Fig. 11 with two A1A and A3A sensors as inputs are compared with the results of multi-class SVM and MLP models trained with these two sensors. It should be noted that for multi-class SVM and MLP models, the RMS values of these two measurements and the correlation value between them are used as input features. Table 12 gives the fault detection results as the classification accuracies for each fault class and the overall accuracy for these three models. As can be seen, the accuracy of multi-class SVM and MLP models are totally decreased compared to the cases when all four axial measurements are used as input (see Table 10). On the other hand, by reducing the number of measurements, effectiveness of the CNN model not only does not drop but improves to full detection accuracy (see Tables 10 and 12). This improvement occurs due to removing the A4A sensor data from input to the CNN model. As stated, data coming from A4A is unsuitable for fault detection of the main bearing, and because of its higher vibration level compared to other measurements, it will reduce the CNN model's efficiency.

5. Conclusion

In this work, a data-driven fault detection method is presented based on the optimized vibration measurements in terms of their location and number for the 5-MW drivetrain system of a spar-type floating wind turbine under different environmental conditions. The studied vibration measurements are chosen according to the suggested measuring positions by ISO standards (ISO 10816–21, for example) for the drivetrain system of the horizontal axis wind turbines with a gearbox. The reference gearbox model simulated in SIMPACK is used in order to get the numerical acceleration responses in radial and axial directions. The faults are modeled in a set of bearings inside the drivetrain MBS model by the change in their stiffness values.

As the first step, correlation analysis was conducted to investigate the relationship between the measurements under different fault and environmental conditions. This study provides an overview of whether and which measurements can be implemented in the proposed fault detection method based on the degree of their correlation. The results showed that the axial measurements obtained from the coaxial shafts (main, low-speed, and intermediate-speed shafts) are highly correlated compared to the other measurement obtained from the high-speed shaft. Also, the correlation values are changed based on the location of the faulty bearings. So that the correlation of the acceleration measurement close to the fault location generally decreases with the other ones.

In the following, a data-driven fault detection method using a deep CNN model was implemented to investigate the detection accuracies in the case of different measurements. Raw axial and radial measurements

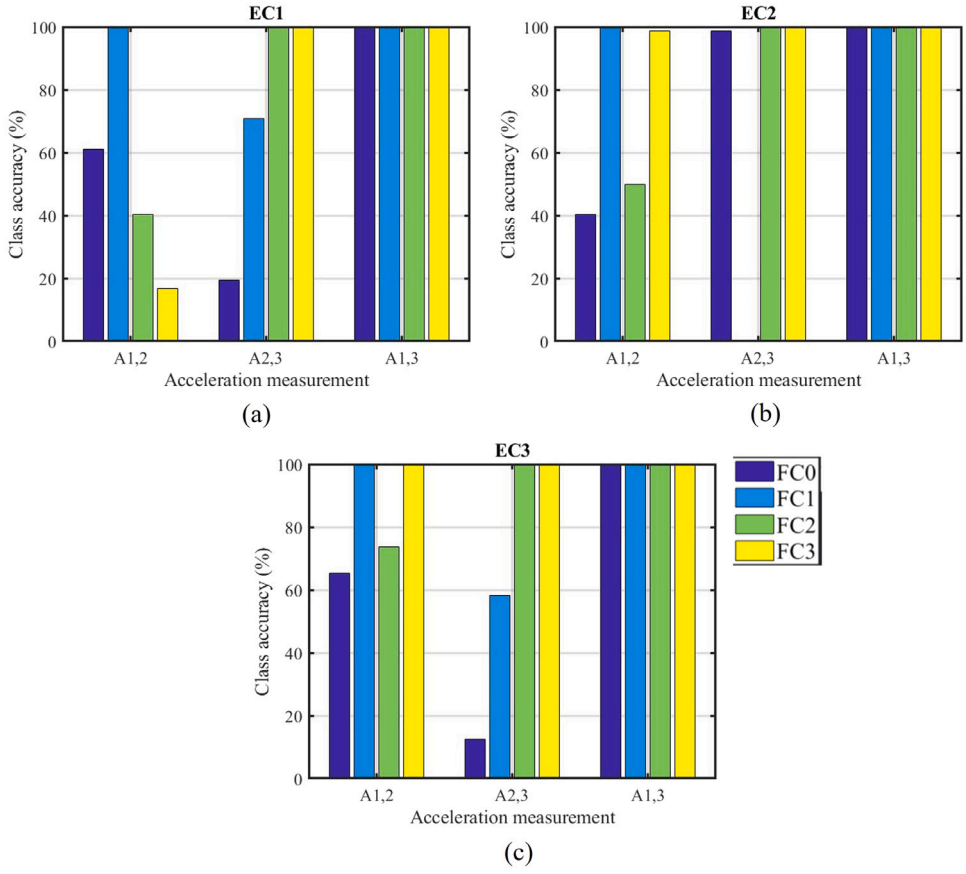


Fig. 11. CNN fault classification results with possible combinations of the correlated axial measurements (a) EC1, (b) EC2, and (c) EC3.

Table 12
Fault classification results for three different machine learning models with optimized A1A and A3A measurements as input.

Environmental condition	ML model	Input feature	FC0 (%)	FC1 (%)	FC2 (%)	FC3 (%)	Overall (%)
EC1	SVM	RMS + ρ	38.9	30.6	95.9	43.1	52.1
	MLP	RMS + ρ	45.4	49.3	93.4	52.2	60.1
	CNN	–	100	100	100	100	100
EC2	SVM	RMS + ρ	75	95.9	100	93.1	91
	MLP	RMS + ρ	81.7	97.4	100	95.6	93.7
	CNN	–	100	100	100	100	100
EC3	SVM	RMS + ρ	43.1	25	100	55.6	56
	MLP	RMS + ρ	34.3	50	100	46.5	57.7
	CNN	–	100	100	100	100	100

were fed directly into the CNN model for the first step. It was shown that the axial measurements give higher total detection accuracy than the radial ones for different fault conditions. Radial measurements show weakness in detecting the main bearing defect, resulting in misclassification between the healthy and main bearing fault classes. PCA is applied to raw measurements before feeding to CNN for the next step. PCA is used as a dimension reduction tool to sort the original data space based on the data variance along each independent eigenvector or principal component. As a result, one could keep the maximum information of all correlated raw measurements with the first few principal components. Because the radial data were not correlated, PCA was ineffective, and each measurement was distributed along with one of the independent principal components. On the other hand, the first two principal components were chosen as the input to CNN for the axial data. It was shown that the training and convergence speed of the CNN model would improve by using PCA and only one of the principal components.

Based on the PCA-CNN fault detection results, it was also observed that using only the correlated axial acceleration measurements, all fault conditions can be detected with 100% accuracy. The high-speed shaft axial vibration, uncorrelated with the others, disrupts the detection accuracy. This measurement cannot effectively detect the main bearing fault because the first and second-stage bearings mainly capture the additional loads caused by this failure. Among the correlated measurements, the two ones from the main and intermediate-speed shafts are enough in order to give 100% detection accuracy for all three faulty bearings. The axial acceleration from the main shaft side quite well detects the main bearing fault. Also, the intermediate-speed shaft acceleration can fully detect the other two fault conditions because of its quite central location on the drivetrain. Any change in the system is well displayed in the vibration level of this measurement. As a result, based on this two-sided axial sensor design and the presented data-driven method, a reliable fault detection approach is proposed for the 5-MW wind turbine drivetrain. The main contribution of this work is that the data-driven method is efficiently adapted to the 5-MW drivetrain using the optimal selection of vibration measurements. Also, this approach can be a promising method for fault detection using the

virtual digital twin model of the drivetrain, where enough vibration data can be acquired from any desired position of the model.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors gratefully acknowledge the financial support of the Research Council of Norway through the InteDiag-WTCP project (Project Number 309205).

References

[1] J. Lee, F. Zhao, Annual-Wind-Report-2022_screen_final_April, 2022.
[2] Global Wind Energy Council, Global offshore wind: Annual market report 2020, in: Global Offshore Wind Report 2020 February, 2020, p. 130.
[3] Seongpil Cho, Zhen Gao, Torgeir Moan, Model-based fault detection, fault isolation and fault-tolerant control of a blade pitch system in floating wind turbines, Renew. Energy (ISSN: 18790682) 120 (2018) 306–321, http://dx.doi.org/10.1016/j.renene.2017.12.102.
[4] Pablo Hevia-Koch, Henrik Klinge Jacobsen, Comparing offshore and onshore wind development considering acceptance costs, Energy Policy (ISSN: 0301-4215) 125 (2019) 9–19, http://dx.doi.org/10.1016/J.ENPOL.2018.10.019.
[5] Zhiwei Gao, Xiaoxu Liu, An overview on fault diagnosis, prognosis and resilient control for wind turbine systems, Processes (ISSN: 22279717) 9 (2) (2021) 1–19, http://dx.doi.org/10.3390/pr9020300.
[6] J. Izquierdo, A. Crespo Márquez, J. Uribetxebarria, A. Erguido, On the importance of assessing the operational context impact on maintenance management for life cycle cost of wind energy projects, Renew. Energy (ISSN: 18790682) 153 (2020) 1100–1110, http://dx.doi.org/10.1016/j.renene.2020.02.048.
[7] Amir Rasekhi Nejad, Modelling and analysis of drivetrains in offshore wind turbines, in: Anaya-Lara O., et al. (Eds.), Offshore Wind Energy Technology, ISBN: 9781119097761, 2018.
[8] Song Xue, Ian Howard, Torsional vibration signal analysis as a diagnostic tool for planetary gear fault detection, Mech. Syst. Signal Process. (ISSN: 10961216) (2018) http://dx.doi.org/10.1016/j.ymssp.2017.07.038.
[9] Zhipeng Feng, Sifeng Qin, Ming Liang, Time-frequency analysis based on Vold-Kalman filter and higher order energy separation for fault diagnosis of wind turbine planetary gearbox under nonstationary conditions, Renew. Energy (ISSN: 18790682) (2016) http://dx.doi.org/10.1016/j.renene.2015.06.041.
[10] Joel Igba, Kazem Alemzadeh, Christopher Durugbo, Egill Thor Eiriksson, Analysing RMS and peak values of vibration signals for condition monitoring of wind turbine gearboxes, Renew. Energy (ISSN: 18790682) (2016) http://dx.doi.org/10.1016/j.renene.2016.01.006.
[11] Joel Igba, Kazem Alemzadeh, Christopher Durugbo, Keld Henningsen, Performance assessment of wind turbine gearboxes using in-service data: Current approaches and future trends, 2015, http://dx.doi.org/10.1016/j.rser.2015.04.139.
[12] Junda Zhu, Jae M. Yoon, David He, Eric Bechhoefer, Online particle-contaminated lubrication oil condition monitoring and remaining useful life prediction for wind turbines, Wind Energy (ISSN: 10991824) (2015) http://dx.doi.org/10.1002/we.1746.
[13] Amir R. Nejad, Peter Fogh Odgaard, Torgeir Moan, Conceptual study of a gearbox fault detection method applied on a 5-MW spar-type floating wind turbine, Wind Energy (ISSN: 10991824) 21 (11) (2018) 1064–1075, http://dx.doi.org/10.1002/we.2213.
[14] ISO10816-21, Mechanical vibration, evaluation of machine vibration by measurements on non-rotating parts: Horizontal axis wind turbines with gearbox, 2015.
[15] ISO16079-2, Condition monitoring and diagnostics of wind turbines: Monitoring the drivetrain, 2020.
[16] Tangbo Bai, Jianwei Yang, Lixiang Duan, Yanxue Wang, Fault diagnosis method research of mechanical equipment based on sensor correlation analysis and deep learning, Shock Vib. (ISSN: 10709622) 2020 (2020) http://dx.doi.org/10.1155/2020/8898944.
[17] Bin Sen Peng, Hong Xia, Yong Kuo Liu, Bo Yang, Dan Guo, Shao Min Zhu, Research on intelligent fault diagnosis method for nuclear power plant based on correlation analysis and deep belief network, Prog. Nucl. Energy (ISSN: 01491970) 108 (May) (2018) 419–427, http://dx.doi.org/10.1016/j.pnucene.2018.06.003.
[18] D. Zhu, X. Yi, Y. Wang, K. Sabra, Structural damage detection through cross correlation analysis of mobile sensing data, in: Proc. of the 5th World Conference, 2010, pp. 1–10, URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.727.1768&rep=rep1&type=pdf.
[19] Jianbin Xiong, Qiong Liang, Jiafu Wan, Qinghua Zhang, Xuehua Chen, Rubao Ma, The order statistics correlation coefficient and PPMCC fuse non-dimension in fault diagnosis of rotating petrochemical unit, IEEE Sens. J. (ISSN: 1530437X) 18 (11) (2018) 4704–4714, http://dx.doi.org/10.1109/JSEN.2018.2820170.
[20] Samira Zare, Moosa Ayati, Simultaneous fault diagnosis of wind turbine using multichannel convolutional neural networks, ISA Trans. (ISSN: 00190578) 108 (2021) 230–239, http://dx.doi.org/10.1016/j.isatra.2020.08.021.
[21] Yanting Li, Wenbo Jiang, Guangyao Zhang, Lianjie Shu, Wind turbine fault diagnosis based on transfer learning and convolutional autoencoder with small-scale data, Renew. Energy (ISSN: 18790682) 171 (2021) 103–115, http://dx.doi.org/10.1016/j.renene.2021.01.143.
[22] Xiaoxu Liu, Zhiwei Gao, Michael Z.Q. Chen, Takagi-sugeno fuzzy model based fault estimation and signal compensation with application to wind turbines, IEEE Trans. Ind. Electron. (ISSN: 02780046) 64 (7) (2017) 5678–5689, http://dx.doi.org/10.1109/TIE.2017.2677327.
[23] Xiaoan Yan, Ying Liu, Yadong Xu, Minping Jia, Multichannel fault diagnosis of wind turbine driving system using multivariate singular spectrum decomposition and improved Kolmogorov complexity, Renew. Energy (ISSN: 18790682) 170 (2021) 724–748, http://dx.doi.org/10.1016/j.renene.2021.02.011.
[24] Jinhao Lei, Chao Liu, Dongxiang Jiang, Fault diagnosis of wind turbine based on long short-term memory networks, Renew. Energy (ISSN: 18790682) 133 (2019) 422–432, http://dx.doi.org/10.1016/j.renene.2018.10.031.
[25] Yue Cui, Pramod Bangalore, Lina Bertling Tjernberg, A fault detection framework using recurrent neural networks for condition monitoring of wind turbines, Wind Energy (ISSN: 1099-1824) 24 (11) (2021) 1249–1262, http://dx.doi.org/10.1002/WE.2628, URL: https://onlinelibrary.wiley.com/doi/full/10.1002/we.2628 https://onlinelibrary.wiley.com/doi/abs/10.1002/we.2628 https://onlinelibrary.wiley.com/doi/10.1002/we.2628.
[26] Ali Dibaj, Amir R. Nejad, Zhen Gao, A data-driven approach for fault diagnosis of drivetrain system in a spar-type floating wind turbine based on the multi-point acceleration measurements, J. Phys. Conf. Ser. (ISSN: 1742-6596) 2265 (3) (2022) 032096, http://dx.doi.org/10.1088/1742-6596/2265/3/032096, URL: https://iopscience.iop.org/article/10.1088/1742-6596/2265/3/032096 https://iopscience.iop.org/article/10.1088/1742-6596/2265/3/032096/meta.
[27] Amir Rasekhi Nejad, Yi Guo, Zhen Gao, Torgeir Moan, Development of a 5 MW reference gearbox for offshore wind turbines, Wind Energy (ISSN: 10991824) (2016) http://dx.doi.org/10.1002/we.1884.
[28] Walt Musial, Sandy Butterfield, Brian Mcniff, Improving wind turbine gearbox reliability, in: European Wind Energy Conference and Exhibition 2007, EWEC 2007, ISBN: 9781622764686, 2007.
[29] Harald Ormberg, Erin E. Bachynski, Global analysis of floating wind turbines: Code development, model sensitivity and benchmark study, in: Proceedings of the International Offshore and Polar Engineering Conference, ISBN: 9781880653944, 2012.
[30] Jm Jonkman, S. Butterfield, W. Musial, G. Scott, Definition of a 5-MW reference wind turbine for offshore system development, Contract (ISSN: 01487299) (2009).
[31] J. Jonkman, Definition of the floating system for phase IV of OC3, Contract (2010).
[32] Simpack MBS Software | Dassault Systèmes. URL: https://www.3ds.com/products-services/simulia/products/simpack/.
[33] Amir Rasekhi Nejad, Erin E. Bachynski, Marit I. Kvittem, Chenyu Luan, Zhen Gao, Torgeir Moan, Stochastic dynamic load effect and fatigue damage analysis of drivetrains in land-based and TLP, spar and semi-submersible floating wind turbines, Mar. Struct. (ISSN: 09518339) 42 (7491) (2015) 137–153, http://dx.doi.org/10.1016/j.marstruc.2015.03.006.
[34] Anil Dhanola, H.C. Garg, Tribological challenges and advancements in wind turbine bearings: A review, Eng. Fail. Anal. (ISSN: 1350-6307) 118 (2020) 104885, http://dx.doi.org/10.1016/J.ENGFAILANAL.2020.104885.
[35] Yi Guo, Robert G. Parker, Stiffness matrix calculation of rolling element bearings using a finite element/contact mechanics model, Mech. Mach. Theory (ISSN: 0094114X) (2012) http://dx.doi.org/10.1016/j.mechmachtheory.2011.12.006.
[36] Aydin Gunduz, Rajendra Singh, Stiffness matrix formulation for double row angular contact ball bearings: Analytical development and validation, J. Sound Vib. (ISSN: 0022460X) 332 (22) (2013) 5898–5916, http://dx.doi.org/10.1016/j.jsv.2013.04.049.
[37] Amir R. Nejad, Torgeir Moan, On model-based system approach for health monitoring of drivetrains in floating wind turbines, Procedia Eng. (ISSN: 18777058) 199 (2017) 2202–2207, http://dx.doi.org/10.1016/j.proeng.2017.09.182.
[38] Steven T. Garren, Maximum likelihood estimation of the correlation coefficient in a bivariate normal model with missing data, Statist. Probab. Lett. (ISSN: 01677152) 38 (3) (1998) 281–288, http://dx.doi.org/10.1016/s0167-7152(98)00035-2.
[39] I.T. Jolliffe, Principal Component Analysis, second ed., ISBN: 0387954422, 2002, http://dx.doi.org/10.2307/1270093.
[40] Ali Dibaj, Mir Mohammad Ettefagh, Reza Hassannejad, Mir Biuok Ehghaghi, A hybrid fine-tuned VMD and CNN scheme for untrained compound fault diagnosis of rotating machinery with unequal-severity faults, Expert Syst. Appl. (ISSN: 09574174) (2020) 114094, http://dx.doi.org/10.1016/j.eswa.2020.114094, URL: https://linkinghub.elsevier.com/retrieve/pii/S0957417420308496.

[41] Matthew D. Zeiler, Rob Fergus, Stochastic pooling for regularization of deep convolutional neural networks, 2013, pp. 1–9, URL: http://arxiv.org/abs/1301.3557.

[42] Jay Lee, Fangji Wu, Wenyu Zhao, Masoud Ghaffari, Linxia Liao, David Siegel, Prognostics and health management design for rotary machinery systems - Reviews, methodology and applications, Mech. Syst. Signal Process. (ISSN: 08883270) 42 (1–2) (2014) 314–334, http://dx.doi.org/10.1016/j.ymssp.2013.06.004.

[43] Young Jun Yoo, Data-driven fault detection process using correlation based clustering, Comput. Ind. (ISSN: 01663615) 122 (2020) http://dx.doi.org/10.1016/j.compind.2020.103279.

[44] Yun Lai Zhou, Hongyou Cao, Quanmin Liu, Magd Abdel Wahab, Output-based structural damage detection by using correlation analysis together with transmissibility, Materials (ISSN: 19961944) 10 (8) (2017) 1–17, http://dx.doi.org/10.3390/ma10080866.

[45] Amir R. Nejad, Erin E. Bachynski, Torgeir Moan, Effect of axial acceleration on drivetrain responses in a spar-type floating wind turbine, J. Offshore Mech. Arct. Eng. (ISSN: 1528896X) 141 (3) (2019) 1–7, http://dx.doi.org/10.1115/1.4041996.

[46] A.C.C. Tan, Y.H. Kim, V. Kosse, Condition monitoring of low-speed bearings - A review, Aust. J. Mech. Eng. (ISSN: 14484846) 6 (1) (2008) 61–68, http://dx.doi.org/10.1080/14484846.2008.11464558.