



Article

Wind Power Converter Fault Diagnosis Using Reduced Kernel PCA-Based BiLSTM

Khadija Attouri ¹, Majdi Mansouri ^{2,*} , Mansour Hajji ¹, Abdelmalek Kouadri ³, Kais Bouzrara ⁴ 
and Hazem Nounou ²

¹ Research Unit Advanced Materials and Nanotechnologies (UR16ES03), Higher Institute of Applied Sciences and Technology of Kasserine, Kairouan University, Kasserine 1200, Tunisia

² Electrical and Computer Engineering Program, Texas A&M University at Qatar, Doha P.O. Box 23874, Qatar

³ Signals and Systems Laboratory, Institute of Electrical and Electronic Engineering, University M'Hamed Bougara of Boumerdes, Avenue of Independence, Boumerdes 35000, Algeria

⁴ Laboratory of Automatic Signal and Image Processing, National Engineering School of Monastir, Monastir 5035, Tunisia

* Correspondence: majdi.mansouri@qatar.tamu.edu

Abstract: In this paper, we present a novel and effective fault detection and diagnosis (FDD) method for a wind energy converter (WEC) system with a nominal power of 15 KW, which is designed to significantly reduce the complexity and computation time and possibly increase the accuracy of fault diagnosis. This strategy involves three significant steps: first, a size reduction procedure is applied to the training dataset, which uses hierarchical K-means clustering and Euclidean distance schemes; second, both significantly reduced training datasets are utilized by the KPCA technique to extract and select the most sensitive and significant features; and finally, in order to distinguish between the diverse WEC system operating modes, the selected features are used to train a bidirectional long-short-term memory classifier (BiLSTM). In this study, various fault scenarios (short-circuit (SC) faults and open-circuit (OC) faults) were injected, and each scenario comprised different cases (simple, multiple, and mixed faults) on different sides and locations (generator-side converter and grid-side converter) to ensure a comprehensive and global evaluation. The obtained results show that the proposed strategy for FDD via both applied dataset size reduction methods not only improves the accuracy but also provides an efficient reduction in computation time and storage space.

Keywords: wind energy converter (WEC) systems; fault detection and diagnosis (FDD); dataset reduction; kernel principal component analysis (KPCA); bidirectional long-short-term memory (BiLSTM)



Citation: Attouri, K.; Mansouri, M.; Hajji, M.; Kouadri, A.; Bouzrara, K.; Nounou, H. Wind Power Converter Fault Diagnosis Using Reduced Kernel PCA-Based BiLSTM. *Sustainability* **2023**, *15*, 3191. <https://doi.org/10.3390/su15043191>

Academic Editor: Hua Li

Received: 15 December 2022

Revised: 10 January 2023

Accepted: 19 January 2023

Published: 9 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Wind energy has been one of the most promising and popular sources and has shown a maximization trend for renewable energy during the past few decades. Wind energy systems (WES) produced more than 420 GW in 2016; by the 2030s, this is expected to grow to more than 1000 GW [1]. However, in this system, failures can nearly always occur anywhere and are categorized as electrical faults, including faulty generators, stator winding short circuits, converter failure, transformer overheating, etc.; electronic faults, which occur frequently in sensors and in electronic boards; and mechanical faults related to the gearbox and the blades. Due to this, it is challenging to ensure the stability, safety, efficiency, performance, and reliability of such systems. Many techniques and methods are applied for the condition monitoring of wind turbine systems. For instance, the research on a particular transient method appropriate for mechanical and electrical fault diagnosis in an induction generator-based wind turbine is provided in [2]. Additionally, to deal with specific types of faults and their detection and classification, a new technique based on deep learning through time-series analysis and convolutional neural networks (CNNs) was duly used [3]. In the same way, an intelligent FDD system (IFDDS) based on developed adaptive

residual convolutional neural networks (ARCNNs) is adopted for small modular reactors (SMRs) [4]. In addition, a generalized ANN method for numerous and diverse wind turbines, for early FDD in the main shaft rear bearing, was developed [5]. Thus, To enhance the classification task using an ANN classifier, an effective feature selection approach based on an improved extension of particle swarm optimization (PSO) was established via suitably reduced datasets of WEC systems [6]. In addition, an ANN-based ensemble classifier was built with the use of bagging, boosting, and random subspace combination techniques and by means of computational and storage cost minimization [7]. On the other hand, for automated fault detection and condition monitoring of wind turbines operating under different statuses, the Wilcoxon rank sum test is properly considered [8]. Further, a reduced Gaussian process regression-based random forest method was utilized to identify and diagnose failures that can occur in a nonlinear WEC system [9].

Therefore, for linearly connected datasets, various data-driven fault diagnostic techniques have been created and essentially evaluated. The PCA, which is used frequently for dimensionality reduction and industrial process monitoring, is the most widely utilized technique [10]. In spite of this, the PCA technique's proven effectiveness is predicated on the assumption that the system is linear and that it can only evaluate linear connections between process variables. This concept may not be applicable to many complicated industrial processes with nonlinearities [11]. Several nonlinear PCA methods have been proposed in the literature to address the aforementioned issues. For instance, a generalized Lorenz model with coexisting attractors was analyzed using a kernel PCA (KPCA) [12]. Moreover, a novel fault detection algorithm for the dynamic nonlinear processes was developed [13].

KPCA, which was first proposed by Scholkopf [14], is the most popular technique that attracts a lot of researchers. It converts input data from the original space to a high-dimensional feature space, where the statistical indices Q and T^2 are extracted and then used for fault detection in the same way as PCA. Thus, the KPCA technique has a great ability to capture nonlinear relationships between variables, providing better results and monitoring a variety of industrial processes [15], but when the training data are very numerous, the computational and storage costs become very high. To overcome these drawbacks, numerous reduced KPCAs (RKPCAs) have been investigated [16–20]. A heuristic K-means clustering algorithm based on the kernel PCA and dynamic programming was developed [21]. Clustering for high-dimensional, low-sample-size (HDLSS) data has been considered [22]. The authors explained the selection of the scale parameter that results in good performance of the KPCA with the Gaussian kernel and provided a theoretical justification for why the Gaussian kernel is useful for clustering high-dimensional data. Then, they used microarray datasets to evaluate the effectiveness and performance of the clustering procedure. A reduced KPCA technique based on K-means clustering was presented, which seeks to find a reduced dataset among the training data in the input space and uses this reduced data to build the KPCA model in the feature space [23].

For the purpose of detecting and diagnosing faults in WEC systems, an improved RNN approach via reduced observations based on HK-means clustering was developed [24]. Thus, an intelligent fault diagnosis method based on reduced kernel PCA-based bidirectional long short-time memory (RKPCA-based BiLSTM) has been established. The contributions of this study are threefold: First, size reduction based on Euclidean distance as a dissimilarity metric between samples and hierarchical K-means clustering methods was performed to deal with the cases of redundancies and to extract reduced observations from the training dataset. Second, the obtained dataset was fed to a kernel PCA technique to extract and select the more pertinent features from the WEC systems. Then, the selected features were used by the BiLSTM classifier, which has been proven to accelerate model convergence, improve classification performance, and enhance the model's ability to extract context information, thereby further improving the accuracy of fault diagnosis, reducing time complexity, and reducing energy consumption. In the following experiments, we compared the BiLSTM classifier with several machine learning and deep learning classifiers—ANN, MNN, CFNN, FFNN, GRNN, PNN, RNN, LSTM, and CNN—to prove

the high efficiency and performance of the developed algorithm. Six arbitrary groups based on the mean, variance, skewness, kurtosis, retained KPCs, Q , and combined index ϕ of features were properly used for fault classification. The remainder of this paper is organized as follows: In Section 2, the RKPCA-based BiLSTM diagnosis algorithm is introduced in detail, including the concepts of LSTM, along with BiLSTM algorithms and a brief mathematical description of the KPCA and reduced KPCA models. The proposed method was tested on wind power converter systems, and the obtained results are summarized in Section 3; we focus on the effects of dataset size reduction and diagnosis accuracy to ensure the performance of our algorithm while reducing computational and storage costs. Section 4 lists some conclusions.

2. Reduced KPCA-Based BiLSTM Algorithm

2.1. Concept of LSTM

Long short-term memory (LSTM) is an artificial recurrent neural network (RNN) architecture that can process entire data sequences in addition to single data points. Figure 1 depicts the LSTM cell's structure.

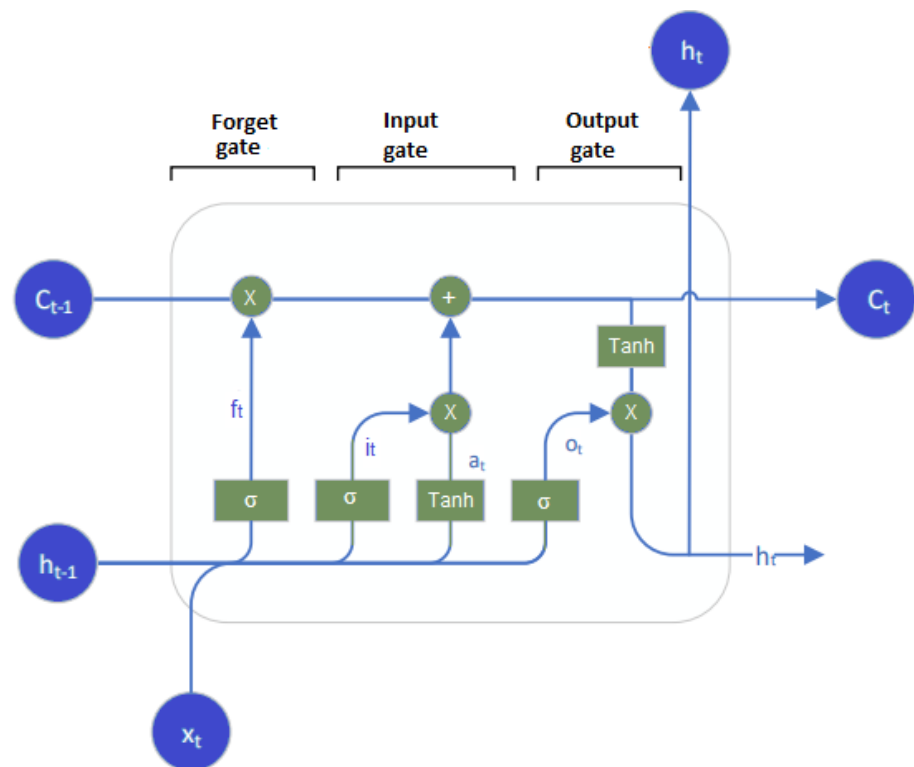


Figure 1. LSTM architecture.

Three gates are present in each cell of the LSTM networks, as shown in Figure 1, and their mathematical descriptions are as follows. The following equation indicates the forget gate, which can control the cell to eliminate certain information from the previous hidden state C_{t-1} :

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (1)$$

where f_t stands for the forget gate's output, h_{t-1} for the preview's output, x_t for the current input vector, and σ for the sigmoid function. The f_t matrix's elements are all included within the range $[0, 1]$, where zero denotes a complete dropout and one denotes a full reservation. The following equation determines the input gate, which controls the cell to reserve a part of the information for the hidden state:

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (2)$$

$$\tilde{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c) \quad (3)$$

where the input gate's output is represented by \tilde{C}_t . When we decide what to reserve and what to forget, the state of the cell will be updated as follows:

$$C_t = f_t \odot C_{t-1} \odot i_t \odot \tilde{C}_t \quad (4)$$

where the symbol \odot indicates the element-wise vectors' multiplication and C_t denotes the long-term state. The output gate, which determines the current cell's output, is provided by:

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o), \text{ and} \quad (5)$$

$$h_t = o_t \odot \tanh(C_t) \quad (6)$$

where h_t represents the output.

2.2. Concept of BiLSTM

The BiLSTM network's structure is depicted in Figure 2. It is clear that the neural network's two LSTM layers are not simply stacked. The output is created by combining the calculations from the two layers after they have separately scanned the data in two directions [25].

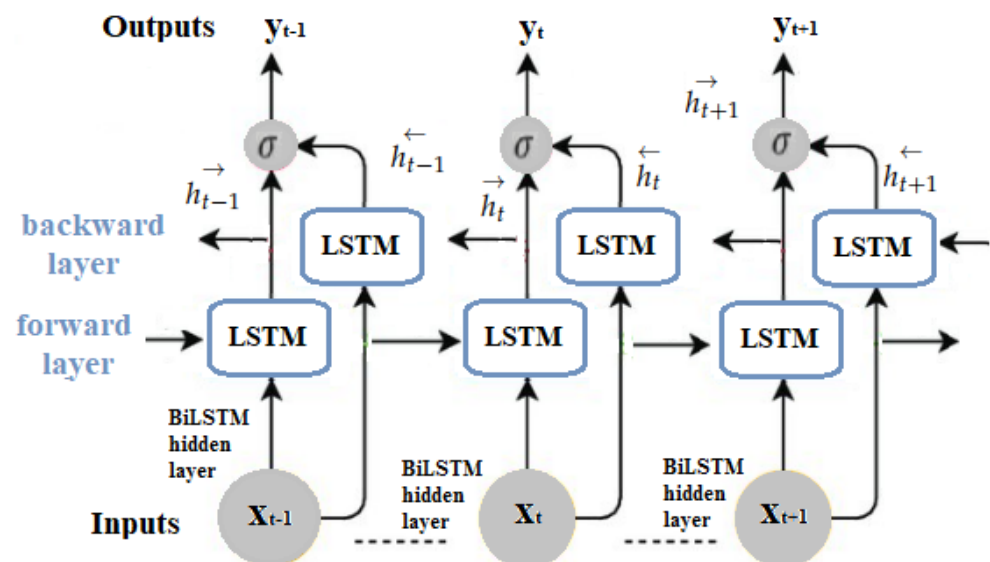


Figure 2. Bidirectional LSTM model.

The BiLSTM algorithm has the ability to learn inputs in two directions: the forward direction (from left to right) and the backward direction (from right to left), and their hidden states are illustrated, respectively, as follows:

$$\vec{h}_t = LSTM(x_t, \vec{h}_{t-1}) \quad (7)$$

$$\overleftarrow{h}_t = LSTM(x_t, \overleftarrow{h}_{t+1}) \quad (8)$$

The BiLSTM output is generated by the concatenation of the forward state (from left to right) and the backward state (from right to left):

$$h_t = [\vec{h}_t, \overleftarrow{h}_t] \quad (9)$$

The final hidden state h_f encodes the majority of the features from the input signal and is employed as an input to the fully connected layer, which aims to transform it into a vector whose length is equal to the class number. The classification of faults is approved using a softmax layer. The probability distribution can be calculated using the following equation:

$$\tilde{Y} = \text{softmax}(W_z h_f + b_z) \quad (10)$$

with W_z denoting the weight and b_z denoting the bias.

$$\text{softmax}(g_i) = \frac{\exp(g_i)}{\sum_{j=1}^k \exp(g_j)} \quad (11)$$

g_i denotes the i th element of the input vector g . By minimizing the error between the predicted \tilde{Y} and actual Y , the BiLSTM model is trained.

2.3. The Concept of the KPCA Model

KPCA is a nonlinear extension of the PCA technique [26], which is one of the most widely used data analysis and dimensional reduction techniques. Therefore, the major goal of kernel PCA is to overcome the limitation of PCA, which only considers the variation in a linear relationship. To preserve the nonlinear structure when we apply PCA to our data, we need to use the kernel trick. The main concept behind KPCA is to use nonlinear mapping to map data into a feature space, and then a linear PCA is performed in that space. Consider the following set of normalized training data: $X = [x(1) \ x(2) \ \dots \ x(n)]^T \in \mathbb{R}^{n \times m}$, where n denotes the number of observations and m denotes the number of process variables. Therefore, the training dataset is mapped into a high-dimensional feature space using a nonlinear mapping in the feature space H , $\phi : x_i \in \mathbb{R}^m \rightarrow \phi_i = \phi(x_i) \in \mathbb{R}^h$, where $h \gg m$ is the dimension in that space [27].

The following formula can be used to calculate the dot product of two vectors, $\phi(x_i)$ and $\phi(x_j)$, with $i, j = 1, \dots, n$:

$$\phi(x_i)^T \phi(x_j) = k(x_i, x_j) \quad (12)$$

where the kernel function is denoted by k . One of the most commonly used kernel functions is the radial basis function (RBF), which is given by the following equation:

$$k(x_i, x_j) = \exp \left[\frac{-\|x_i - x_j\|^2}{2\delta^2} \right] \quad (13)$$

where the width of a Gaussian function that controls the kernel's flexibility is denoted by the symbol δ . A typical selection for δ is the average minimum distance (d) between two points in the training dataset, i.e., $\delta^2 = c \frac{1}{n} \sum_{i=1}^n \min_{j \neq i} d^2(x_i, x_j)$, as suggested in [28], where c is a variable that the user defines.

Supposing that the vectors in the feature space are scaled to zero mean and unit variance, the mapped data are arranged as $\chi = [\phi(x_1) \ \phi(x_2) \ \dots \ \phi(x_n)]^T$. The dataset's covariance matrix C in the feature space can be calculated using the following equation:

$$\begin{aligned} (n-1)C &= \chi^T \chi \\ &= \sum_{i=1}^n \phi_i \phi_i^T \end{aligned} \quad (14)$$

KPCA in the feature space is equivalent to solving the eigenvector equation shown below:

$$\begin{aligned}\chi^T \chi v s. &= \sum_{i=1}^n \phi_i \phi_i^T v \\ &= \lambda v s.\end{aligned}\quad (15)$$

The kernel function $k(x_i, x_j) = \phi(x_i)^T \phi(x_j)$ can be used to evaluate the Gram matrix $\chi^T \chi$, even though the mapping function ϕ is not explicitly defined. The matrix K can be defined with $k(x_i, x_j)$ elements using the kernel trick.

$$\begin{aligned}K = \chi \chi^T &= \begin{bmatrix} \phi_1^T \phi_1 & \dots & \phi_1^T \phi_n \\ \vdots & \dots & \vdots \\ \phi_n^T \phi_1 & \dots & \phi_n^T \phi_n \end{bmatrix} \\ &= \begin{bmatrix} k(x_1, x_1) & \dots & k(x_1, x_n) \\ \vdots & \dots & \vdots \\ k(x_n, x_1) & \dots & k(x_n, x_n) \end{bmatrix}\end{aligned}\quad (16)$$

The eigenvector equation in the feature space is resolved by KPCA. Let λ be the corresponding eigenvalue of the eigenvector α of the matrix K .

$$v = \lambda^{-1} \chi^T \alpha \quad (17)$$

The matrix of the KPCA's retained principal loading in the feature space is represented by $P = [v_1, \dots, v_l] \in \mathfrak{R}^{n \times l}$, and the $n - l$ last principal loading is denoted by $\tilde{P} = [v_{l+1}, \dots, v_n] \in \mathfrak{R}^{n \times (n-l)}$.

$$P = \left[\frac{1}{\lambda_1} \chi^T \alpha_1, \dots, \frac{1}{\lambda_l} \chi^T \alpha_l \right] \quad (18)$$

The selection of the PCs number has been the focus of numerous studies, some of which are described in [29,30]. The scores are calculated by the use of the below equations for a measurement x and its mapped vector $\phi = \phi(x)$:

$$t = P^T \phi \in \mathfrak{R}^l \quad (19)$$

$$\tilde{t} = \tilde{P}^T \phi \in \mathfrak{R}^{n-l} \quad (20)$$

2.4. The Concept of the Reduced KPCA Model

Since training data used for modeling must be stored and used for monitoring as well, KPCA has memory and computational difficulties. When there are many observations, especially when monitoring dynamic processes, a difficulty arises. Recently, several solutions for a time-varying system have been suggested in order to develop a reduced KPCA (RKPCA) model using a reduced training dataset, among other potential techniques. They consist of obtaining a reduced set of kernel vectors using, for example, k-means clustering, hierarchical clustering, PCA, Euclidean distance (ED), and so on.

2.4.1. Reduced KPCA Based on Euclidean Distance

When two or more observations are redundant, the Euclidean distance (ED) is employed as a dissimilarity metric to keep only one observation. Additionally, among the m measurement variables, we extract the most relevant data information using the suggested reduced KPCA (RKPCA) approach. The KPCA model is constructed using the retained

observations as a new data matrix [31]. On the other hand, the computed dissimilarity matrix D for all pairs of samples whose elements are $d_{ij}, i, j = 1 \dots n$ is provided by:

$$D = \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1n} \\ d_{21} & d_{22} & \dots & d_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ d_{n1} & d_{n2} & \dots & d_{nn} \end{bmatrix} \quad (21)$$

for a given data matrix X with n samples and m process variables, where d_{ij} represents the Euclidean distance between the rows x_i and x_j of the data matrix X . Thus, d_{ij} is given by:

$$d_{ij} = \sqrt{\sum_{k=1}^m (x_{ik} - x_{jk})^2} \quad (22)$$

D is a symmetric matrix as a result, and its diagonal elements are null. All measurement redundancy in the original matrix is eliminated based on the dissimilarity distance. This in turn contributes significantly to reducing the number of measurements.

2.4.2. Reduced KPCA Based on Hierarchical K-means Clustering

The hierarchical K-means (HK-means) clustering strategy was developed to reduce the number of training samples. Therefore, the K-means clustering metric is used to establish reduced data of new centers or vectors to accurately represent the original data [32]. Moreover, when using the K-means approach, each observation is assigned to the cluster with the closest distance from the centroid cluster. K-means clustering is aimed at improving the quality of the cluster results given by agglomerative hierarchical clustering. When the number of samples is large, the HK-means approach is employed to reduce the computation time. The advantages of the hierarchical and K-means are combined in this method [33]. The number of clusters is initially determined using the hierarchical technique. Then, to improve the classes and optimize them, it employs the latter for K-means clustering. Additionally, the cluster analysis utilizes a clustering algorithm. It aims to establish a cluster hierarchy that is typically represented as a tree diagram called a dendrogram [33]. Divisive and agglomerative hierarchical clustering are the two different types of hierarchical clustering. A single cluster is given to every observation in the divisive hierarchical technique. The cluster is then divided into two same clusters. Finally, after each cluster is recursively performed, a cluster is defined for each sample. In the agglomerative hierarchical technique, each observation is initially treated as a single-element cluster. Therefore, a new aggregate cluster is created by joining the two nearby clusters. Therefore, this procedure is repeated until all observations are portions of a single large cluster. Different metrics can be used to calculate the distance between two clusters. The Ward's linkage distance is employed in our work. Thus, by the use of the agglomerative hierarchical clustering, the N_r clusters are obtained: $\{c_1', c_2', \dots, c_{N_r}'\}$; $x_j \in c_i'$; $j = 1 \dots, n_i'$, and $i = 1, \dots, N_r$, where n_i' denotes the number of samples in c_i' . The aim of the K-means clustering is to partition the N_r clusters into N_r disjoint clusters. The clusters are partitioned up so that the squared error between a cluster's empirical mean and its points is as small as possible.

$$E = \sum_{i=1}^{N_r} \sum_{x_j \in c_i} \|x_j - M_i\|^2 \quad (23)$$

The following equation defines the calculation of centroid M_i of cluster c_i :

$$M_i = \sum_{x_j \in c_i} \frac{x_j}{n_i} \quad (24)$$

By the use of the HK-means, the input dataset is calculated as follows:

$$X_r = [x_1' \ x_2' \ \dots \ x_{N_r}']^T \quad (25)$$

with

$$x_i' = \frac{1}{n_i} \sum_{j \in c_j} x_j, \quad i = 1, \dots, N_r \quad (26)$$

with $x_j \in R^m$, $j = 1, \dots, n_i$ and $N_r = l + 1$

The following equation defines the mapping of X_r .

$$X_r = [\phi(x_1') \ \phi(x_2') \ \dots \ \phi(x_{N_r}')]^T \in \mathfrak{R}^{N_r \times h} \quad (27)$$

The obtained reduced kernel $K_r \in \mathfrak{R}^{N_r \times N_r}$ is obtained:

$$\begin{aligned} K_r &= X_r (X_r)^T \\ &= \begin{bmatrix} k(x_1', x_1') & \dots & k(x_1', x_{N_r}') \\ \vdots & & \vdots \\ k(x_{N_r}', x_1') & \dots & k(x_{N_r}', x_{N_r}') \end{bmatrix} \end{aligned} \quad (28)$$

The the eigenvalue α_r and the eigenvector λ_r of the new reduced kernel matrix are depicted below:

$$\lambda' \alpha' = k_r \alpha' \quad (29)$$

By extracting and selecting the more pertinent features, the data's dimensional reduction are obtained. Thus, the determination of these relevant features can be calculated as follows:

$$t = \Lambda^{-1/2} P^T k(x) \quad (30)$$

with $P = [\alpha_1', \dots, \alpha_l']$ denoting the l principal eigenvectors of K_r which correspond to the largest eigenvalues $\Lambda' = \text{diag} \{ \lambda_1', \dots, \lambda_l' \}$. Therefore, several methods have been deployed for the selection of the principal components. The cumulative percent variance (CPV) criterion was applied in this study.

2.4.3. Fault Diagnosis Based on the RKPCA-Based BiLSTM Algorithm

The proposed RKPCA-based BiLSTM (RKPCA-BiLSTM) strategy involves three major steps: data size reduction, feature extraction/selection, and fault diagnosis. The main goal of the developed RKPCA-BiLSTM approach is to reduce the complexity and overburden of the BiLSTM classifier, thereby reducing the computation time, which represents an inevitable and crucial challenge in the fault detection and diagnosis domain. Unlike the classical diagnosis techniques, which apply the raw data to the neural network directly, our methodology pre-processed the data and then reduced the training dataset by keeping only the non-redundant, most informative, and pertinent information in order to reduce the computation time, thereby speeding up the convergence of the neural network and may helping to enhance the classification accuracy. In brief, the obtained new data are fed to the KPCA technique after the size reduction stage using Euclidean distance or hierarchical K-means (HK-means) clustering in order to extract the multivariate, statistical, and nonlinear features. Then, the achieved features are fed as input to the BiLSTM classifier in order to detect, classify, and distinguish between the diverse faulty operating modes that may occur in the WEC systems. Figure 3 shows the block diagram that summarizes the various steps of the suggested approach for FDD purposes.

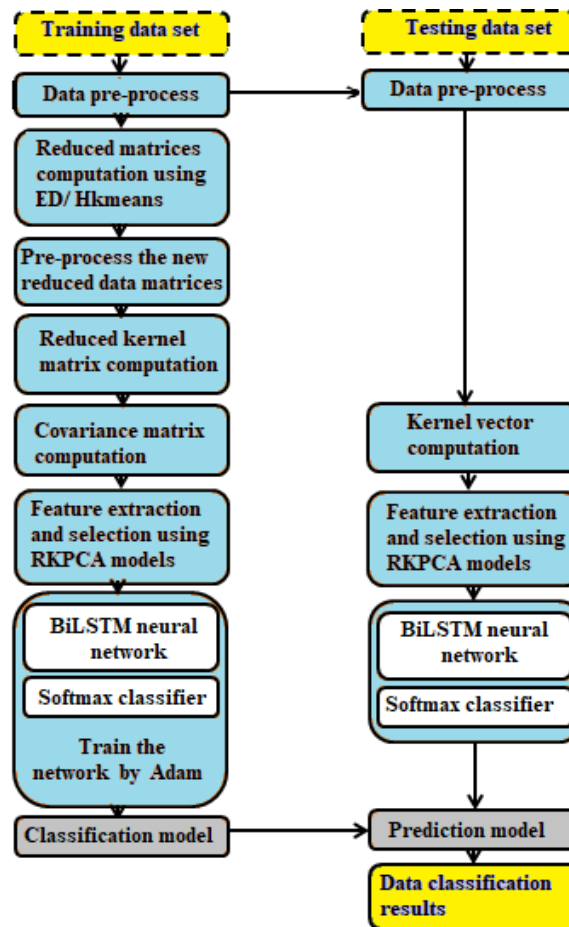


Figure 3. The steps of the proposed approach for fault diagnosis.

The reduced KPCA-BiLSTM algorithm is divided into two phases: the training phase and the testing phase. The detailed descriptions are depicted in the following Algorithm 1:

Algorithm 1 Reduced KPCA-based BiLSTM algorithm.

Input: $n \times m$ data matrix $X_i, i = 1, 2, \dots, n$

Training phase

1. Standardize the training dataset,
2. Reduce the size of the training dataset using ED/HK-means clustering,
3. Map the reduced matrix into the features space,
4. Determine the reduced kernel matrix,
5. Extract and select the more pertinent features using reduced KPCA models,
6. Classify the faults using BiLSTM classifier,
7. Make out the classification model.

Testing phase

1. Standardize the testing dataset using the mean and the variance computed in the training state,
 2. Calculate the kernel vector,
 3. Extract and select the features using the reduced KPCA model,
 4. Classify the faults using the BiLSTM classifier,
 5. Provide the prediction model,
 6. Obtain the fault diagnosis results.
-

2.5. Process Description

A wind turbine system is a complicated electromechanical device that transforms wind energy into electrical energy. This research considers a squirrel cage induction generator (SCIG)-based variable-speed wind turbine, as depicted in Figure 4.

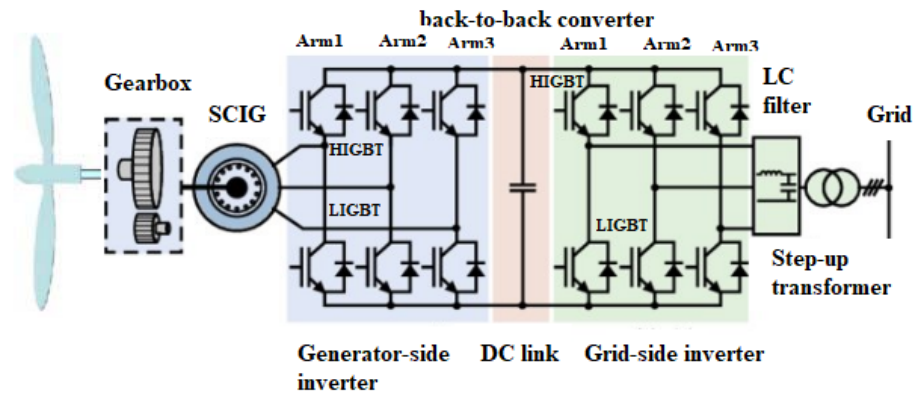


Figure 4. Variable-speed wind turbine based on a SCIG and converter topology.

This structure allows for infinitely variable speed operation. No matter how quickly the machine rotates, the generated voltage is rectified and turned into direct voltage and current. As a result, the control for the grid-side converter helps with providing an alternating voltage with a fixed frequency that refers to the grid. The nominal power of the generator determines the maximum power produced by the turbine. The grid-side generator for this structure is based on an insulated gate bipolar transistor (IGBT), whose structure is identical to the grid-side converter's. Table 1 displays the characteristics of the wind turbines.

Table 1. Parameters for the wind turbine [25].

Parameters	Nomenclature	Values
Nominal Power of turbine	P_{tn}	15 kW
Moment of inertia of turbine	J_t	1000 kgm ²
Stator resistance	R_s	0.087 Ohm
Stator leakage inductance	L_s	0.8 mH
Rotor resistance	R_r	0.228 Ohm
Rotor resistance	R_r	0.228 Ohm
Rotor leakage inductance	L_r	0.8 mH
Magnetizing inductance	L_m	34.7 mH
Number of poles	P	4
Moment of inertia of generator	J_g	0.2 kgm ²

The next section presents the performance evaluation of the developed technique.

3. Simulation and Discussion

3.1. Input Data Description

WEC systems are subject to various faults and failures that can lead to losses in efficiency and performance, ultimately resulting in the system's destruction and damage. According to statistical studies, printed circuit boards (PCB), capacitors, and power semiconductor devices (IGBT) are the major parts of wind power converters that are susceptible to failure. One of the main causes of converter faults is a fault in a power semiconductor device. There are two types of common faults in power semiconductor devices: short-circuit (SC) faults and open-circuit (OC) faults. Power semiconductor device fault mechanisms have been addressed in several literary works, as illustrated in Figure 5.

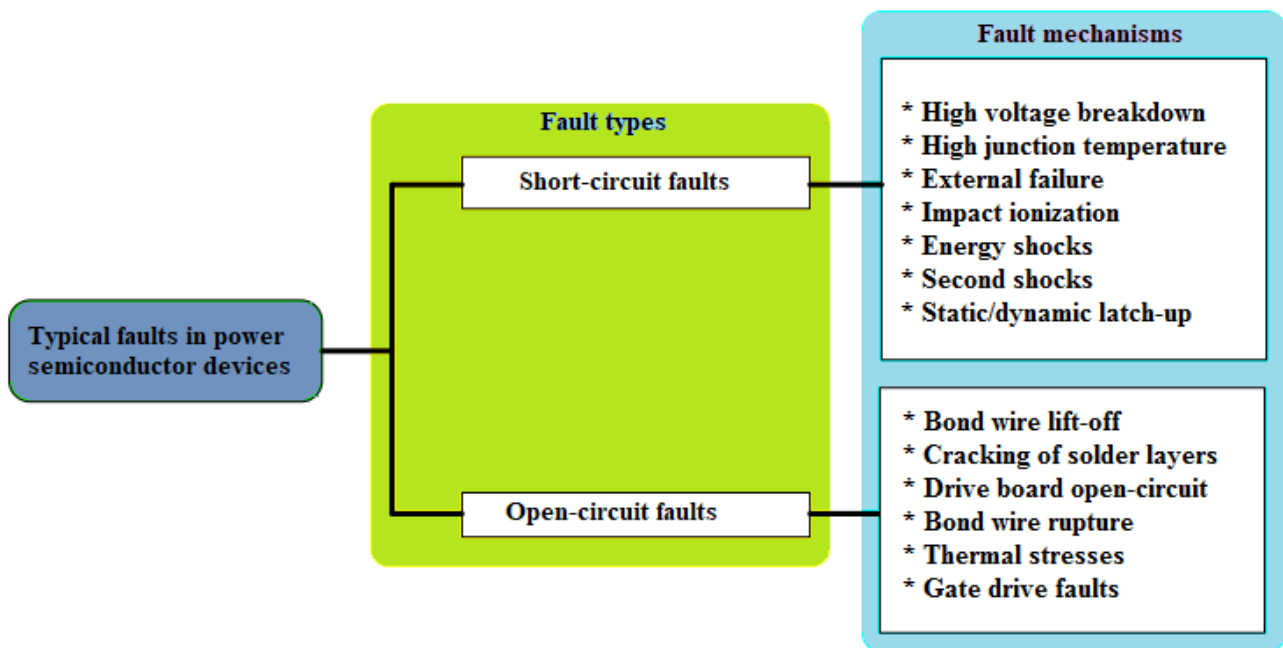


Figure 5. Mechanisms and types of power semiconductor devices.

Power semiconductor device faults in wind power converters are mostly caused by the following: Firstly, the instantaneous voltage or current of power converters can be excessively high when a wind turbine is activated or suffers from powerful gusts. Secondly, heat dissipation performance degradation and fatigue accumulation of a power semiconductor may result in device damage during a lengthy period of operation in wind power converters. Finally, in wind farms, corrosive gases, moisture, and dust can cause abnormal operation or catastrophic defects in power semiconductor devices [34]. For wind power converters, both SC and OC faults in power semiconductor devices can cause significant and serious damage and harm.

To ensure a comprehensive study and global analysis, various fault situations were injected at various locations and on different sides (generator-side converter and grid-side converter), as illustrated in Table 2.

Table 2. Descriptions of the different labeled faults injected into the system.

Fault Side	Type	and	Symbol	Fault Description
Simple fault Generator side			SFGen	Fault1: Short circuit that affects only one High IGBT (HIGBT) in the first Arm of the converter generator side.
				Fault2: Open circuit that affects only one Low IGBT (LIGBT) in the second Arm of the converter generator side.
				Fault3: Open circuit that affects only one High IGBT in the third Arm of the converter generator side.
Simple fault grid side			SFGrid	Fault4: Open circuit that affects only one Low IGBT in the first Arm of the converter grid side.
				Fault5: Short circuit that affects only one High IGBT in the second Arm of the converter grid side.
				Fault6: Short circuit that affects only one Low IGBT in the third Arm of the converter grid side.
Multiple fault generator side			MFGen	Fault7: Short circuit that affects only one Low IGBT in the first Arm of the converter generator side and Open circuit that affects only one High IGBT in the second Arm of the converter generator side.
Multiple fault grid side			MFGrid	Fault8: Short circuit that affects only one Low IGBT in the first Arm of the converter grid side and Open circuit that affects only one High IGBT in the second Arm of the converter grid side.
Mixt fault both side			MxF	Fault9: Short circuit that affects only one Low IGBT in the first Arm of the converter generator side and Open circuit that affects only one High IGBT in the second Arm of the converter grid side.

Ten variables were measured at several locations. The measured variables are illustrated in Table 3.

Table 3. Labeling and descriptions of the measured and monitored system variables.

Variables	Descriptions
x_1	C_m : Mechanical torque (N_m)
x_2	N_g : Generator speed (tr/m)
x_3	i_{sag} : Generator current phase a (A)
x_4	i_{sbg} : Generator current phase b (A)
x_5	i_{scg} : Generator current phase c (A)
x_6	V_{DC} : Bus voltage (V)
x_7	P_{out} : Output power (W)
x_8	i_{sar} : Grid current phase a (A)
x_9	i_{sbr} : Grid current phase b (A)
x_{10}	i_{scr} : Grid current phase c (A)

Figure 6 shows the output power (x_7) behavior for different scenarios, such as the healthy scenario and different faulty operating modes (modes 3 and 9).

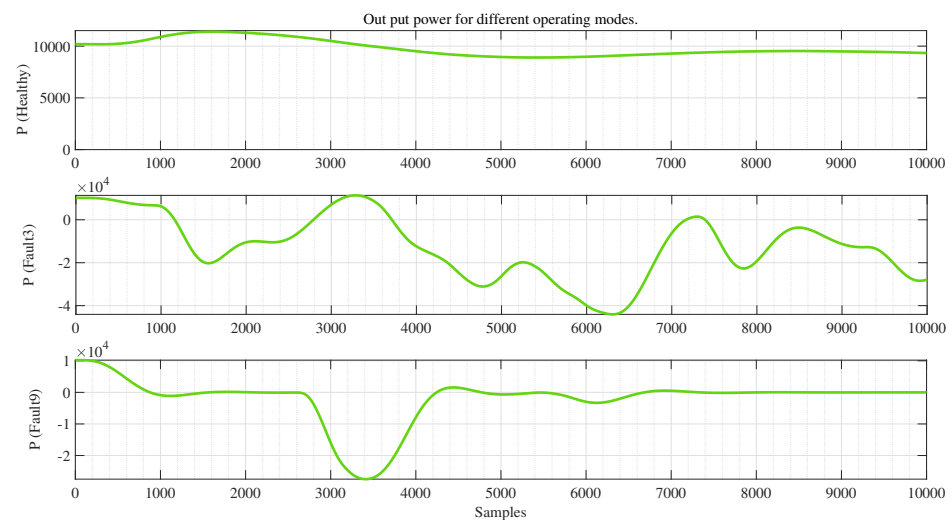


Figure 6. Output power for different scenarios.

It can be seen from Figure 6 that the output power in the healthy case is almost constant (around 10,000 W). However, it is clear that when a fault is injected, as depicted, for instance, in fault 3 and fault 9, the same level of power is found with oscillations, which clearly proves that faults can affect and impact the behavior of the system.

3.2. Fault Diagnosis Results

The operation of the WEC system represents one healthy case which is assigned to a class (C_0) and nine diverse faulty scenarios assigned to (C_1, C_2, \dots , and C_9), as shown in Table 4. In total, 12,500 samples were used to adequately depict each mode's behavior. In the training phase, we employed 80% of the samples, and in the testing phase, 20%.

Table 4. Construction of database for the fault detection and diagnosis system.

Class	State	Training Data	Testing Data
C ₀	Healthy	10,000	2500
C ₁	SFGen	10,000	2500
C ₂		10,000	2500
C ₃		10,000	2500
C ₄	SFGrid	10,000	2500
C ₅		10,000	2500
C ₆		10,000	2500
C ₇	MFGen	10,000	2500
C ₈	MFGGrid	10,000	2500
C ₉	MxF	10,000	2500

Several machine learning and deep learning classifiers were used in this study, and the best one was selected on the basis of its efficiency and classification accuracy (please refer to Table 5). The optimal hyper-parameters employed in the current study are illustrated in Table 6.

To maintain and select the retained KPCs, a 95% cumulative variance criterion was used. Note that 52 and 25 retained KPCs were obtained for the RKPCA_{HK-means} clustering and RKPCA_{ED}, respectively.

Table 5. Performance comparison of classification techniques.

Methods	Global Performance				
	Accuracy	Recall	Precision	F1 Score	CT(s)
ANN	67.98	67.98	68.48	68.22	0.17
MNN	70.13	70.13	71.54	70.82	0.29
CFNN	68.14	58.14	65.81	61.73	0.40
FFNN	64.76	64.76	63.47	64.10	0.37
GRNN	69.79	69.79	70.80	70.29	0.64
PNN	10.72	10.72	11.12	10.91	1.34
RNN	64.68	55.58	66.20	60.42	0.32
CNN	28.19	28.19	28.19	28.19	1.19
LSTM	80.68	82.64	80.99	81.80	1.19
BiLSTM	83.52	83.52	83.52	83.52	1.40
KPCA-BiLSTM	98.80	98.80	98.46	98.62	2.26

The number of hidden layers chosen was 10, and the number of neurons in each hidden layer was 50 for the ANN, MNN, FFNN, CFNN, GRNN, PNN, and RNN classifiers. We used a convolution layer, Relu function, pooling layer, fully connected layer, and softmax layer for the CNN classifier. Additionally, to train the neural network, the CNN employed the cross-entropy loss function and the Adam optimization algorithm. A Matlab environment was used to implement these different classifiers.

It is evident that the KPCA-BiLSTM classifier offers better classification performance. In order to reduce the computation time (2.26 s) and thereby the complexity, and speed up the BiLSTM classifier (and perhaps further enhance and improve the classification accuracy), an efficient FDD-approach is proposed using RKPCA-based BiLSTM, in which the training dataset is reduced using the ED or the HK-means clustering tools, and then the new reduced dataset is fed to the KPCA technique in order to extract and select the most informative features. The extracted features are then introduced to the BiLSTM to classify faults and distinguish between the diverse operating modes.

Table 6. Hyperparameter settings.

Hperparameters	Values
Optimizer	Adam
Loss function	Cross entropy
Dropout	0.2
Learning rate	0.001
Regularizer	L2
MaxEpoches	20
Mini batch size	250
BiLSTM layer nodes	50

Typically, there are two main categories of feature selection strategies: supervised techniques and unsupervised techniques. All of the following are considered as supervised techniques: relief [35], Fisher score [36], Chi-squared score [37], correlation-based feature selection [38], and fast correlation-based filter [39].

The variance [40]; mean [40]; kurtosis [41]; skewness [41]; mean absolute difference [41], dispersion ratio [41]; multicluster feature selection [42]; Laplacian score [43], Laplacian score combined with distance-based entropy [43]; and multivariate and univariate statistical methods, such as the T^2 statistic [44], squared prediction error (SPE) statistic [45], combined index ϕ [46], and generalized likelihood ratio [47], are the most popular unsupervised methods.

In this work, SPE, ϕ , sampled variance, mean, skewness, and kurtosis metrics were used for feature selection. Therefore, in order to achieve a global study, we compare the sampled mean, kurtosis, skewness, and variance metrics; the retained KPCs in terms of efficiency and classification accuracy; and the performance of their combination, in order to study which features are better than others.

The accuracy results show that kurtosis, skewness, mean, and variance are not ideal features; for example, the kurtosis, variance, and skewness accuracy were 11.44%, 10%, and 35.75%, respectively (through the use of $RKPCA_{HK-means}$). In order to increase the efficiency of these features, a combination of them with the ℓ -retained KPCs was applied in this study. The description of the used groups are illustrated in Table 7.

Table 7. Selected features for fault diagnosis.

Groups	Descriptions
group1	SPE and ϕ indices
group2	skewness and the ℓ retained KPC_s
group3	kurtosis and the ℓ retained KPC_s
group4	variance and the ℓ retained KPC_s
group5	mean and the ℓ retained KPC_s
group6	The first ℓ KPC_s

More details on the performance for different groups are presented in Table 8. The obtained number of the reduced samples represents a ratio of 25% of the training dataset, by the use of the $RKPCA_{HK-means}$ and the $RKPCA_{ED}$.

Table 8. Performance comparison of different groups used.

Group Label	Group Description	Global Performance				
		Accuracy	Recall	Precision	F1 Score	CT(s)
group1	RKPCA _{ED}	30	30	30	30	0.70
	RKPCA _{HK-means}	40	40	40	40	0.98
group2	RKPCA _{ED}	71.39	80	71.39	75.45	0.53
	RKPCA _{HK-means}	51.05	51.05	51.05	51.05	0.84
group3	RKPCA _{ED}	61.51	61.51	61.51	61.51	0.96
	RKPCA _{HK-means}	61.97	61.97	61.97	61.97	0.64
group4	RKPCA _{ED}	50	50	50	50	0.95
	RKPCA _{HK-means}	50.24	50.24	50.24	50.24	0.71
group5	RKPCA _{ED}	81.50	81.50	81.50	81.50	0.93
	RKPCA _{HK-means}	51.71	51.71	51.70	51.70	0.96
group6	RKPCA _{ED}	100	100	100	100	0.57
	RKPCA _{HK-means}	100	100	100	100	0.28

As we can see in Table 8, the RKPCA_{ED} and the RKPCA_{HK-means} models were attained and required the best diagnosis results with 100% accuracy by the use of the sixth group (the first ℓ KPCs).

It can be seen from Table 9 that the proposed RKPCA-based-BiLSTM approach clearly reduces the computation time from 2.26 to 0.57 s or 0.28 s through the use of the RKPCA_{ED} or RKPCA_{HK-means} clustering, respectively. This accelerates the BiLSTM classifier and slows down its convergence, and improves and increases the fault classification accuracy by achieving and attaining a perfect result of 100%. The testing classification results using the confusion matrix for RKPCA_{ED} and RKPCA_{HK-means} approaches are the same.

Table 9. Performance comparison of different techniques.

Techniques	Global Performance				
	Accuracy	Recall	Precision	F1 Score	CT(s)
BiLSTM	83.52	83.52	83.52	83.52	1.40
KPCA-BiLSTM	98.80	98.80	98.46	98.62	2.26
RKPCA _{ED} -BiLSTM	100	100	100	100	0.57
RKPCA _{HK-means} -BiLSTM	100	100	100	100	0.28

The confusion matrix represents the misclassified and the correct observations for both healthy and different faulty scenarios. Predicted process statuses and the true classes were determined by the x -axis and y -axis, respectively.

We can see from this confusion matrix (Table 10) that for the healthy (C_0) and faulty cases (C_1, C_2, \dots, C_9), the two suggested methodologies correctly identified 2500 samples from 2500 (true positive), which means that the numerous scenarios are well and correctly classified, and the misclassification rate is 0%.

One can conclude that the suggested methods achieved the best overall performance with an accuracy of 100%. Therefore, the proposed methodologies are considered as perfect alternatives for fault classification due to their high accuracy and reliability.

Table 10. Confusion matrix for RKPCA_{ED} and RKPCA_{HK-means}-clustering-based BiLSTM in testing phase.

True Classes	Predicted Classes										Recall
	C ₀	C ₁	C ₂	C ₃	C ₄	C ₅	C ₆	C ₇	C ₈	C ₉	
C ₀	2500	0	0	0	0	0	0	0	0	0	100
C ₁	0	2500	0	0	0	0	0	0	0	0	100
C ₂	0	0	2500	0	0	0	0	0	0	0	100
C ₃	0	0	0	2500	0	0	0	0	0	0	100
C ₄	0	0	0	0	2500	0	0	0	0	0	100
C ₅	0	0	0	0	0	2500	0	0	0	0	100
C ₆	0	0	0	0	0	0	2500	0	0	0	100
C ₇	0	0	0	0	0	0	0	2500	0	0	100
C ₈	0	0	0	0	0	0	0	0	2500	0	100
C ₉	0	0	0	0	0	0	0	0	0	2500	100
Precision	100	100	100	100	100	100	100	100	100	100	100

4. Conclusions

Efficient fault detection and diagnosis (FDD) methods for wind energy conversion systems were proposed in this paper. Unlike the classical diagnosis techniques, which apply the raw data directly to a neural network, our strategies (RKPCA_{ED}-based BiLSTM and RKPCA_{HK-means}-clustering-based BiLSTM) improve on the traditional BiLSTM and preprocess the data by reducing the number of samples in the data matrix to build the reduced reference model, extracting multivariate and nonlinear features, and selecting the most pertinent and informative characteristics. In order to evaluate the classification performance, the suggested FDD techniques were compared to other traditional methods, including ANN, MNN, CFNN, FNN, GRNN, PNN, CNN, RNN, and LSTM. The recommended methods are utilized in order to reduce the dimensions of data, which can accelerate the neural network's convergence and reduce the computation time (from 2.26 to 0.28 s) and the memory space, and also increase the classification accuracy (100% accuracy).

Author Contributions: Methodology, K.A., M.M., M.H. and A.K.; Validation, K.A. and M.M.; Investigation, H.N.; Resources, K.B.; Writing—original draft, K.A.; Writing—review & editing, M.M. and K.B.; Supervision, M.M., M.H., A.K. and H.N. All authors have read and agreed to the published version of the manuscript.

Funding: Open Access funding provided by the Qatar National Library. The publication is the result of the Qatar National Research Fund (QNRF) research grant.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data availability upon Editor request.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

FDD	Fault Detection and Diagnosis
FES	Feature Extraction and Selection
PCA	Principal Component Analysis
KPCA	Kernel Principal Component Analysis
RKPCA _{ED}	Reduced Kernel Principal Component Analysis via Euclidian distance
RKPCA _{HK-means}	Reduced Kernel Principal Component Analysis via Hierarchical K-means
ANN	Artificial Neural Network
MNN	Multilayer Neural Network

CFNN	Cascade forward Neural Network
NN	Neural Network
RNN	Recurrent NN
FFNN	Feed-Foward NN
MNN	Multiple Layers NN
GRNN	Generalized Regression NN
PNN	Probabilistic Neural Network NN
LSTM	Long Short Term Memory
BiLSTM	Bidirectional Long Short Term Memory
CNN	Convolutional Neural Network
ℓ	Number of retained PCs
CPV	Cumulative Percentage of Variance
CT	Computation Time
CM	Confusion Matrix

References

1. Marugán, A.P.; Márquez, F.P.G.; Perez, J.M.P.; Ruiz-Hernández, D. A survey of artificial neural network in wind energy systems. *Appl. Energy* **2018**, *228*, 1822–1836.
2. Al-Ahmar, E.; Benbouzid, M.; Turri, S. Wind energy conversion systems fault diagnosis using wavelet analysis. *Int. Rev. Electr. Eng.* **2008**, *3*, 646–652.
3. Rahimilarki, R.; Gao, Z.; Jin, N.; Zhang, A. Convolutional neural network fault classification based on time-series analysis for benchmark wind turbine machine. *Renew. Energy* **2022**, *185*, 916–931. [[CrossRef](#)]
4. Yao, Y.; Wang, J.; Xie, M. Adaptive residual CNN-based fault detection and diagnosis system of small modular reactors. *Appl. Soft Comput.* **2022**, *114*, 108064. [[CrossRef](#)]
5. Zhang, Z.Y.; Wang, K.S. Wind turbine fault detection based on SCADA data analysis using ANN. *Adv. Manuf.* **2014**, *2*, 70–78. [[CrossRef](#)]
6. Mansouri, M.; Dhibi, K.; Nounou, H.; Nounou, M. An Effective Fault Diagnosis Technique for Wind Energy Conversion Systems Based on an Improved Particle Swarm Optimization. *Sustainability* **2022**, *14*, 11195.
7. Dhibi, K.; Mansouri, M.; Bouzrara, K.; Nounou, H.; Nounou, M. Reduced neural network based ensemble approach for fault detection and diagnosis of wind energy converter systems. *Renew. Energy* **2022**, *194*, 778–787. [[CrossRef](#)]
8. Dao, P.B. On Wilcoxon rank sum test for condition monitoring and fault detection of wind turbines. *Appl. Energy* **2022**, *318*, 119209. [[CrossRef](#)]
9. Mansouri, M.; Fezai, R.; Trabelsi, M.; Mansour, H.; Nounou, H.; Nounou, M. Enhanced Gaussian Process Regression for Diagnosing Wind Energy Conversion Systems. *IFAC-PapersOnLine* **2022**, *55*, 673–678. [[CrossRef](#)]
10. George, J.P.; Chen, Z.; Shaw, P. Fault detection of drinking water treatment process using PCA and Hotelling's T2 chart. *Int. J. Comput. Inf. Eng.* **2009**, *3*, 430–435.
11. Mika, S.; Schölkopf, B.; Smola, A.; Müller, K.R.; Scholz, M.; Rätsch, G. Kernel PCA and de-noising in feature spaces. *Adv. Neural Inf. Process. Syst.* **1998**, *11*.
12. Cui, J.; Shen, B.W. A kernel principal component analysis of coexisting attractors within a generalized Lorenz model. *Chaos Solitons Fractals* **2021**, *146*, 110865.
13. Zhang, Q.; Li, P.; Lang, X.; Miao, A. Improved dynamic kernel principal component analysis for fault detection. *Measurement* **2020**, *158*, 107738. [[CrossRef](#)]
14. Schölkopf, B.; Smola, A.; Müller, K.R. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comput.* **1998**, *10*, 1299–1319. [[CrossRef](#)]
15. Lee, J.M.; Yoo, C.; Choi, S.W.; Vanrolleghem, P.A.; Lee, I.B. Nonlinear process monitoring using kernel principal component analysis. *Chem. Eng. Sci.* **2004**, *59*, 223–234. [[CrossRef](#)]
16. Kaib, M.T.H.; Kouadri, A.; Harkat, M.F.; Bensmail, A. RKPCA-based approach for fault detection in large scale systems using variogram method. *Chemom. Intell. Lab. Syst.* **2022**, *225*, 104558. [[CrossRef](#)]
17. Bencheikh, F.; Harkat, M.; Kouadri, A.; Bensmail, A. New reduced kernel PCA for fault detection and diagnosis in cement rotary kiln. *Chemom. Intell. Lab. Syst.* **2020**, *204*, 104091. [[CrossRef](#)]
18. Lahdhiri, H.; Elaissi, I.; Taouali, O.; Harkat, M.F.; Messaoud, H. Nonlinear process monitoring based on new reduced Rank-KPCA method. *Stoch. Environ. Res. Risk Assess.* **2018**, *32*, 1833–1848. [[CrossRef](#)]
19. Taouali, O.; Jaffel, I.; Lahdhiri, H.; Harkat, M.F.; Messaoud, H. New fault detection method based on reduced kernel principal component analysis (RKPCA). *Int. J. Adv. Manuf. Technol.* **2016**, *85*, 1547–1552.
20. Jaffel, I.; Taouali, O.; Harkat, M.F.; Messaoud, H. Moving window KPCA with reduced complexity for nonlinear dynamic process monitoring. *ISA Trans.* **2016**, *64*, 184–192. [[CrossRef](#)]
21. Xu, M.; Franti, P. A heuristic K-means clustering algorithm by kernel PCA. In Proceedings of the 2004 International Conference on Image Processing, 2004, ICIP'04, Singapore, 24–27 October 2004; Volume 5, pp. 3503–3506.

22. Nakayama, Y.; Yata, K.; Aoshima, M. Clustering by principal component analysis with Gaussian kernel in high-dimension, low-sample-size settings. *J. Multivar. Anal.* **2021**, *185*, 104779. [[CrossRef](#)]
23. Fezai, R.; Mansouri, M.; Taouali, O.; Harkat, M.F.; Nounou, H. Reduced kernel principal component analysis for fault detection and its application to an air quality monitoring network. In Proceedings of the 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Miyazaki, Japan, 7–10 October 2018; pp. 3159–3164.
24. Mansouri, M.; Dhibi, K.; Hajji, M.; Bouzara, K.; Nounou, H.; Nounou, M. Interval-Valued Reduced RNN for Fault Detection and Diagnosis for Wind Energy Conversion Systems. *IEEE Sens. J.* **2022**, *22*, 13581–13588. [[CrossRef](#)]
25. Yahyaoui, Z.; Hajji, M.; Mansouri, M.; Abodayeh, K.; Bouzara, K.; Nounou, H. Effective Fault Detection and Diagnosis for Power Converters in Wind Turbine Systems Using KPCA-Based BiLSTM. *Energies* **2022**, *15*, 6127. [[CrossRef](#)]
26. Harkat, M.F.; Kouadri, A.; Fezai, R.; Mansouri, M.; Nounou, H.; Nounou, M. Machine learning-based reduced kernel PCA model for nonlinear chemical process monitoring. *J. Control. Autom. Electr. Syst.* **2020**, *31*, 1196–1209. [[CrossRef](#)]
27. Mansouri, M.; Baklouti, R.; Harkat, M.F.; Nounou, M.; Nounou, H.; Hamida, A.B. Kernel generalized likelihood ratio test for fault detection of biological systems. *IEEE Trans. Nanobiosci.* **2018**, *17*, 498–506. [[CrossRef](#)]
28. Rathi, Y.; Dambreville, S.; Tannenbaum, A. Statistical shape analysis using kernel PCA. In Proceedings of the Image Processing: Algorithms and Systems, Neural Networks, and Machine Learning, San Jose, CA, USA, 16–18 January 2006; Volume 6064, pp. 425–432.
29. Valle, S.; Li, W.; Qin, S.J. Selection of the number of principal components: The variance of the reconstruction error criterion with a comparison to other methods. *Ind. Eng. Chem. Res.* **1999**, *38*, 4389–4401. [[CrossRef](#)]
30. Tamura, M.; Tsujita, S. A study on the number of principal components and sensitivity of fault detection using PCA. *Comput. Chem. Eng.* **2007**, *31*, 1035–1046.
31. Dhibi, K.; Fezai, R.; Mansouri, M.; Kouadri, A.; Harkat, M.F.; Bouzara, K.; Nounou, H.; Nounou, M. A hybrid approach for process monitoring: improving data-driven methodologies with dataset size reduction and interval-valued representation. *IEEE Sens. J.* **2020**, *20*, 10228–10239. [[CrossRef](#)]
32. Zhu, J.; Jiang, Z.; Evangelidis, G.D.; Zhang, C.; Pang, S.; Li, Z. Efficient registration of multi-view point sets by K-means clustering. *Inf. Sci.* **2019**, *488*, 205–218. [[CrossRef](#)]
33. Nguyen, H.; Bui, X.N.; Tran, Q.H.; Mai, N.L. A new soft computing model for estimating and controlling blast-produced ground vibration based on hierarchical K-means clustering and cubist algorithms. *Appl. Soft Comput.* **2019**, *77*, 376–386. [[CrossRef](#)]
34. Liang, J.; Zhang, K.; Al-Durra, A.; Muyeen, S.; Zhou, D. A state-of-the-art review on wind power converter fault diagnosis. *Energy Rep.* **2022**, *8*, 5341–5369. [[CrossRef](#)]
35. Pérez-Ortiz, M.; Torres-Jiménez, M.; Gutiérrez, P.A.; Sánchez-Monedero, J.; Hervás-Martínez, C. Fisher score-based feature selection for ordinal classification: A social survey on subjective well-being. In Proceedings of the International Conference on Hybrid Artificial Intelligence Systems, Seville, Spain, 18–20 April 2016; pp. 597–608.
36. Thaseen, I.S.; Kumar, C.A. Intrusion detection model using fusion of chi-square feature selection and multi class SVM. *J. King Saud Univ.-Comput. Inf. Sci.* **2017**, *29*, 462–472.
37. Karegowda, A.G.; Manjunath, A.; Jayaram, M. Comparative study of attribute selection using gain ratio and correlation based feature selection. *Int. J. Inf. Technol. Knowl. Manag.* **2010**, *2*, 271–277.
38. Yu, L.; Liu, H. Feature selection for high-dimensional data: A fast correlation-based filter solution. In Proceedings of the 20th International Conference on Machine Learning (ICML-03), Washington, DC, USA, 21–24 August 2003; pp. 856–863.
39. Yang, C.; Zhang, W.; Zou, J.; Hu, S.; Qiu, J. Feature selection in decision systems: A mean-variance approach. *Math. Probl. Eng.* **2013**, *2013*, 268063.
40. Doraisamy, S.; Golzari, S.; Mohd, N.; Sulaiman, M.N.; Udzir, N.I. A Study on Feature Selection and Classification Techniques for Automatic Genre Classification of Traditional Malay Music. In Proceedings of the ISMIR, Philadelphia, PA, USA, 14–18 September 2008; pp. 331–336.
41. Wagner, J.; Kim, J.; André, E. From physiological signals to emotions: Implementing and comparing selected methods for feature extraction and classification. In Proceedings of the 2005 IEEE International Conference on Multimedia and Expo, Amsterdam, The Netherlands, 6–8 July 2005; pp. 940–943.
42. Liu, R.; Yang, N.; Ding, X.; Ma, L. An unsupervised feature selection algorithm: Laplacian score combined with distance-based entropy measure. In Proceedings of the 2009 Third International Symposium on Intelligent Information Technology Application, NanChang, China, 21–22 November 2009; Volume 3, pp. 65–68.
43. He, X.; Cai, D.; Niyogi, P. Laplacian score for feature selection. *Adv. Neural Inf. Process. Syst.* **2005**, *18*.
44. Song, F.; Guo, Z.; Mei, D. Feature selection using principal component analysis. In Proceedings of the 2010 International Conference on System Science, Engineering Design and Manufacturing Informatization, Yichang, China, 12–14 November 2010; Volume 1, pp. 27–30.
45. Sheriff, M.Z.; Botre, C.; Mansouri, M.; Nounou, H.; Nounou, M.; Karim, M.N. Process monitoring using data-based fault detection techniques: Comparative studies. *Fault Diagn. Detect.* **2017**, *32*, 137–144.

46. Mansouri, M.; Harkat, M.F.; Nounou, H.; Nounou, M.N. *Data-Driven and Model-Based Methods for Fault Detection and Diagnosis*; Elsevier: Amsterdam, The Netherlands, 2020.
47. Mansouri, M.; Hajji, M.; Trabelsi, M.; Harkat, M.F.; Al-khazraji, A.; Livera, A.; Nounou, H.; Nounou, M. An effective statistical fault detection technique for grid connected photovoltaic systems based on an improved generalized likelihood ratio test. *Energy* **2018**, *159*, 842–856. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.