IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received November 9, 2020, accepted November 30, 2020, date of publication December 2, 2020,
date of current version December 17, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3042101

A Novel Fault Diagnosis of Uncertain Systems Based on Interval Gaussian Process Regression: Application to Wind Energy Conversion Systems

MAJDI MANSOURI[1], (Member, IEEE), RADHIA FEZAI[2], MOHAMED TRABELSI[3], (Senior Member, IEEE), MANSOUR HAJJI[4], MOHAMED-FAOUZI HARKAT[5], (Member, IEEE), HAZEM NOUNOU[1], (Senior Member, IEEE), MOHAMED N. NOUNOU[6], (Senior Member, IEEE), AND KAIS BOUZRARA[2]

[1]Electrical and Computer Engineering Program, Texas A&M University at Qatar, Doha, Qatar
[2]Research Laboratory of Automation, Signal Processing and Image, National Engineering School of Monastir, Monastir 5019, Tunisia
[3]Department of Electronic and Communications Engineering, Kuwait College of Science and Technology, Kuwait 13133, Kuwait
[4]Higher Institute of Applied Science and Technology of Kasserine, University of Kairouan, Kairouan 1200, Tunisia
[5]Department of Electronics, Badji Mokhtar–Annaba University, Annaba 23000, Algeria
[6]Chemical Engineering Program, Texas A&M University at Qatar, Doha, Qatar

Corresponding author: Majdi Mansouri (majdi.mansouri@qatar.tamu.edu)

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

ABSTRACT Fault detection and diagnosis (FDD) of wind energy conversion (WEC) systems play an important role in reducing the maintenance and operational costs and increase system reliability. Thus, this paper proposes a novel Interval Gaussian Process Regression (IGPR)-based Random Forest (RF) technique (IGPR-RF) for diagnosing uncertain WEC systems. In the proposed IGPR-RF technique, the effective interval-valued nonlinear statistical features are extracted and selected using the IGPR model and then fed to the RF algorithm for fault classification purposes. The proposed technique is characterized by a better handling of WEC system uncertainties such as wind variability, noise, measurement errors, which leads to an improved fault classification accuracy. The obtained results show that the proposed IGPR-RF technique is characterized by a high diagnosis accuracy (an average accuracy of 99.99%) compared to the conventional classifiers.

INDEX TERMS Gaussian process regression (GPR), interval-valued data, random forest (RF), feature extraction and selection, fault detection and diagnosis (FDD), wind energy conversion (WEC) systems.

I. INTRODUCTION

The deployment of Wind Energy Conversion (WEC) systems has witnessed an increasing need for the reduction of maintenance and operational costs [1], [2], where the most effective solutions are found in condition monitoring and diagnosis [3]. Indeed, the operation of WEC systems is usually accompanied by unexpected faults, which should be detected and classified at an early stage to avoid a system collapse. The wind variability, vibrations, and mainly the power electronics interfaces remain the main sources of failures [4], [5].

Many fault detection and diagnosis (FDD) approaches have been proposed for WEC systems in the literature [6], [7]. Generally, FDD techniques can be categorized into two main classes: data-driven [8], [9] and model-based techniques [10], [11]. The Data-driven FDD techniques make only use of the available diagnosis data [8], [9]. The data are first applied to build a model in the training phase, which is then applied in the testing phase for diagnosis purposes.

On the other hand, model-based FDD techniques consist in comparing systems' measurements with system variables calculated from the mathematical model, which is usually computed using some fundamental understanding of the system under normal operating conditions [10], [12], [13]. The residual which presents the difference between the measurements and the predicted model can be used as a chart for fault diagnosis.

In [1], [3], the authors presented a brief description of different kinds of faults, their generated signatures, and diagnosis solutions. Using the gearbox vibration signal, the authors

The associate editor coordinating the review of this manuscript and approving it for publication was Yuan Zhuang.

in [14] have proposed a deep learning technique while a multiscale convolutional neural network was proposed in [15] to extract the faulty wind turbine features under different operating modes. In [16], the authors proposed fault detection and identification approaches which can identify faults, determine the occurring time and location, and estimate its severity. The authors in [17] proposed observer-based FDD techniques for wind turbines, where the diagnosed residuals are generated using Kalman filter, the detection phase is addressed using generalized likelihood ratio test, and the isolation phase is achieved using dual sensor redundancy. Finally, the performance of the proposed FDD techniques is assessed using Monte Carlo schemes. In [18], the authors developed a data-driven FDD approach for the gearbox of a WEC system. Moreover, in the paper [19], the authors proposed unknown input observer based scheme for detecting faults in a wind turbine converter. In [20], the authors proposed a data-driven multimode FDD technique to discriminate the WEC system faults. In the developed technique, the wind turbine nonlinear characteristics were approximated by multiple piece-wise linear systems.

Furthermore, several approaches have been developed to improve the overall performance of WEC systems [21]–[23]. The first phase in the WEC system diagnosis is the extraction of the most relevant patterns/features from the original dataset. Gaussian Process Regression (GPR) is one of the most well-known feature extraction and modeling strategies. In [24], it has been shown that the GPR presented an improved modeling and prediction accuracy when compared to the classical techniques. However, the mostly used GPR based diagnosis technique considers only single-valued data and does not take into account the system uncertainties.

To address the above issue, this paper proposes an interval GPR (IGPR) algorithm where the data is interval-valued-represented. The developed IGPR is characterized by a better handling of WEC system uncertainties such as wind variability, noise, measurement errors, which leads to an improved fault classification accuracy.

The IGPR method is applied to extract the multivariate and interval-valued features, including the interval mean vector M_{IGPR} and the interval variance matrix C_{IGPR}. It is characterized by its efficient extraction of multivariate and uncertain patterns from any data set. The developed approach, the so-called IGPR$_{CR}$, consists of concatenating center and range matrices to compute the new numerical matrix and then fitting a GPR model on the matrix.

The interval-valued statistical parameters obtained from the IGPR model, including the interval mean vector M_{IGPR} and the interval variance matrix C_{IGPR}, are then selected as features and fed to the RF classifier for decision making. Indeed, the RF classifier, a combination of tree predictors, has been recently presented as one of the most effective classification techniques in FDD problems [25], [26].

Therefore, the main contribution of the current work is to develop a feature extraction and selection method using IGPR then introduce the selected interval-valued multivariate features to several RF algorithms for classification purposes.

To summarize, the developed approach consists of two phases. First, the IGPR model is applied to the original data in order to extract and select the most accurate features (including the mean vector M_{IGPR} and the variance matrix C_{IGPR}). Then, the M_{IGPR} and C_{IGPR} are introduced to the RF classifier to perform the detection and classification of faults. The main difference between the proposed solution and the conventional RF algorithm is the introduction of a phase that performs features extraction and selection from the entire data. Two kinds of classifiers are considered in this work: a multi-class classifier and a set of one class classifiers The multi-class classifier consists of classifying instances into one or more classes. To better improve the diagnosis abilities, a bank of one-class classifiers is proposed. To illustrate the feasibility and effectiveness of the proposed technique, a WEC system is used as a validation platform. The open-circuit, wear-out, and short-circuit are the three transistor faults considered in this paper. Besides, a comparative study between the proposed technique and other machine learning (ML)-based classifiers including interval kernel PCA-based RF [26], Support Vector Machines (SVM) [27], Decision Tree (DT) [28], Naive Bayes (NB) [29], Discriminant Analysis (DA) [30] and K-Nearest Neighbors (KNN) [31], is presented.

The performance of the proposed techniques is investigated using sets of emulated data extracted under different operating conditions. The presented results confirm the high-effectiveness of the developed technique in monitoring uncertain WEC systems due to the high diagnosis capabilities of the interval-valued features-based IGPR and its ability to distinguish between the different operating modes of the WEC system.

The rest of the paper is structured as follows. Section 2 describes the proposed IGPR-RF technique. The diagnosis results are evaluated using the WEC system data in Section 3. The interpretations and conclusions are drawn in Section 4.

II. DESCRIPTION OF THE PROPOSED METHODOLOGY

A. IGPR FOR FEATURE EXTRACTION AND SELECTION

Nonlinear GPR is a machine learning technique based on Bayesian theory and statistical learning theory. The main idea of GPR is to assume that the learning sample follows the prior probabilities of the Gaussian process and then determines the corresponding posterior probability. It is suitable for complex regression problems such as nonlinear and high dimensionality. However, GPR is used for single-valued data, which is a result of simplification during the data mining procedure. Thus, GPR based on interval-valued data representation is required to describe the data uncertainty and variability.

Assuming that $[x] = [\underline{x}, \overline{x}]$ represents the input interval-valued data unit, where $\underline{x}$ and $\overline{x} \in \mathbf{R}$ and $\underline{x} \leq \overline{x}$. $\underline{x}$ and $\overline{x}$ are called the lower and upper boundary respectively. $[y] = [\underline{y}, \overline{y}]$

represents the output interval-valued data unit, where $\underline{y}$ and $\overline{y}$ are called the lower and upper output boundary, respectively. Denote $[X] = [x_{ij}]$ as an $(N \times m)$ input and output interval-valued matrix as per:

$$[X] = \begin{pmatrix} [\underline{x}_{11}, \overline{x}_{11}] & \cdot & \cdot & [\underline{x}_{1m}, \overline{x}_{1m}] \\ \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot \\ [\underline{x}_{N1}, \overline{x}_{N1}] & \cdot & \cdot & [\underline{x}_{Nm}, \overline{x}_{Nm}] \end{pmatrix} \tag{1}$$

The output interval-valued matrix $[Y] = [y_{ij}] \in \mathbf{R}^{N\times p}$ is defined by:

$$[Y] = \begin{pmatrix} \left[\underline{y}_{11}, \overline{y}_{11}\right] & \cdot & \cdot & \left[\underline{y}_{1p}, \overline{y}_{1p}\right] \\ \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot \\ \left[\underline{y}_{N1}, \overline{y}_{N1}\right] & \cdot & \cdot & \left[\underline{y}_{Np}, \overline{y}_{Np}\right] \end{pmatrix} \tag{2}$$

B. IGPR BASED CENTERS AND RANGES

The basic idea of the proposed IGPR-based Centers and ranges method (IGPR$_{CR}$) is to fitting a GPR model to interval-valued data using the information contained in the centers and ranges of the intervals in order to improve the model prediction performance compared to the classical GPR technique.

The proposed IGPR$_{CR}$ model consists first of transforming the input $[X]$ and output $[Y]$ matrices into numerical matrices based on the interval centers and ranges. The input center X^c and range X^r matrices, and output center Y^c and range Y^r matrices are defined by:

$$\begin{cases} X^c = [x_1^c, x_2^c, \ldots, x_N^c]^T \in \mathbf{R}^{N\times m} \\ X^r = [x_1^r, x_2^r, \ldots, x_N^r]^T \in \mathbf{R}^{N\times m} \\ Y^c = [y_1^c, y_2^c, \ldots, y_N^c]^T \in \mathbf{R}^{N\times q} \\ Y^r = [y_1^r, y_2^r, \ldots, y_N^r]^T \in \mathbf{R}^{N\times p} \end{cases} \tag{3}$$

where the input center x_i^c and range x_i^r vectors, and the ouput center y_i^c and range y_i^r vectors are defined, respectively, by:

$$x_i^c = \frac{1}{2}(\underline{y}_i + \overline{y}_i), \quad x_i^r = \frac{1}{2}(\overline{x}_i) - \underline{x}_i \tag{4}$$

$$y_i^c = \frac{1}{2}(\underline{y}_i + \overline{y}_i), \quad y_i^r = \frac{1}{2}(\overline{y}_i - \underline{y}_i) \tag{5}$$

The new input X_{CR} and output Y_{CR} data matrices are constructed by the concatenation of centers and range data matrices as:

$$\begin{cases} X_{CR} = [X^c \; X^r] \in \mathcal{R}^{N\times 2m} \\ Y_{CR} = [Y^c \; Y^r] \in \mathcal{R}^{N\times 2p} \end{cases} \tag{6}$$

For an input vector $x^{CR} = [x^c, x^r]$ and its corresponding output vector $y^{CR} = [y^c, y^r]$, an interval Gaussian process $f(x^{CR})$ can be fully specified by its mean function $m(x^{CR})$ and covariance function $k(x^{CR}, x'^{CR})$. The interval Gaussian process is defined as:

$$f(x^{CR}) = \mathcal{GP}\left(m(x^{CR}), k(x^{CR}, x'^{CR})\right) \tag{7}$$

where $m(x^{CR}) = \mathbf{E}\left[f(x^{CR})\right]$ and $k(x^{CR}, x'^{CR}) = \mathbf{E}\left[\left(f((x^{CR}) - m(x^{CR})\right)\left(f((x'^{CR}) - m(x'^{CR})\right)\right]$.

The covariance function $k(x^{CR}, x'^{CR})$ or the kernel plays an important role in the IGPR operation. A large variety of kernel functions can be used depending on the specific application. In this study, a Gaussian kernel function was chosen for the GPR, which takes the following form [32]:

$$k(x^{CR}, x'^{CR}) = exp(-\frac{1}{2\delta^2} \parallel x^{CR} - x'^{CR} \parallel^2) \tag{8}$$

where δ is the characteristic length-scale. The output vector y^{CR} can be related to an underlying arbitrary regression function $f(x^{CR})$ with an additive independent identically distributed Gaussian noise ϵ, which represents the noise component from the interval data. This relationship is expressed by:

$$y^{CR} = f(x^{CR}) + \epsilon \tag{9}$$

where ϵ is the additive white noise and assumed to be the independent and identically distributed Gaussian noise such that $\epsilon \sim \mathcal{N}(0, \sigma_n^2)$, with σ_n^2 is the standard deviation of this noise.

The interval Gaussian process represented in equation 10 becomes,

$$y^{CR} = \mathcal{GP}\left(m(x^{CR}), k(x^{CR}, x'^{CR} + \sigma_n^2)\right) \tag{10}$$

The prior joint distribution of the observation value Y^{CR} and the prediction value y_*^{CR} can be obtained by:

$$\begin{bmatrix} Y^{CR} \\ y_*^{CR} \end{bmatrix} \sim \mathcal{N}\left(m\begin{bmatrix} X^{CR} \\ x_*^{CR} \end{bmatrix}, \begin{bmatrix} K + \sigma_n^2 I & k_*^T \\ k_* & k_{**} \end{bmatrix}\right) \tag{11}$$

where I is the identity matrix, K is the Gram matrix of training dataset, $k_* = \left[k(x_1^{CR}, x_*^{CR}) \cdots k(x_N^{CR}, x_*^{CR})\right]^T$ and $k_{**} = k(x_*^{CR}, x_*^{CR})$.

Conditioning the joint Gaussian prior distribution based on X^{CR}, Y, and x_*^{CR}, the predictive distribution can be calculated by:

$$p(y_*^{CR} \mid X^{CR}, Y, x_*^{CR}) \sim \mathcal{N}(\overline{y_*}^{CR}, C_{IGPR}) \tag{12}$$

where M_{IGPR} is the predictive mean and C_{IGPR} is the predictive variance which are given respectively, by

$$M_{IGPR} = m(x_*^{CR}) + k_*\left[K + \sigma_n^2 I\right]^{-1}(Y^{CR} - m(X^{CR})) \tag{13}$$

$$C_{IGPR} = k_{**} - k_*\left[K + \sigma_n^2 I\right]^{-1} k_*^T \tag{14}$$

The choice of the M_{IGPR} and C_{IGPR} as input features to the RF classifier should enhance the diagnosis performance. In the following, more details on the methodology are presented.

C. RANDOM FOREST FOR FAULT CLASSIFICATION

Once the statistical quantities M_{IGPR} and C_{IGPR} are computed using the IGPR model, the system faults should be isolated. In the current paper, the RF algorithm will be applied to isolate/classify these faults and distinguish between the different

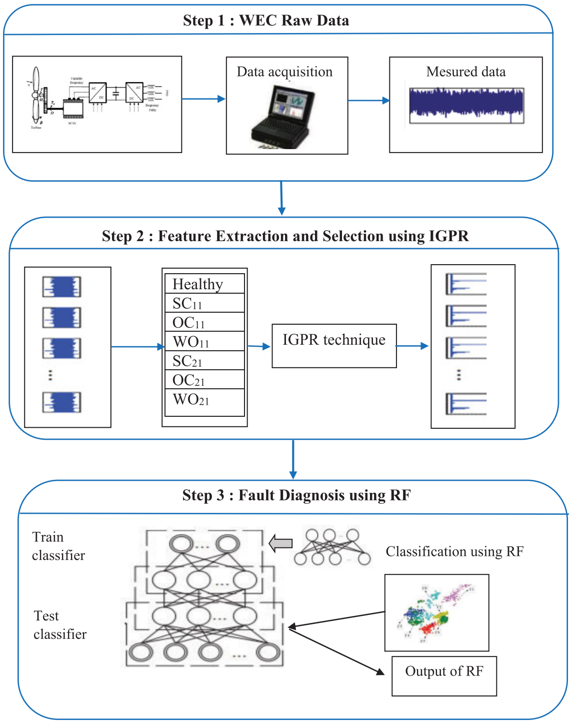

FIGURE 1. Flowchart of the proposed FDD technique.

operating modes. The RF classifier algorithm was developed by Breiman [33] based on the bagging idea. It combines multiple decision trees to create a forest [34], [35]. The features of each generated tree are randomly chosen and then the most popular class is voted. The output of the classifier is obtained by a majority vote of the trees in the forest. The RF classifier is one of the most prevalent algorithms adopted to address the problems of multi-classification. However, the RF implementation suffers from certain drawbacks when considering the correlations between variables. In addition, to perform diagnosis, the RF uses only the raw data by the direct use of measured variables, which might lead to a low performance due to the data redundancies and noises. Therefore, to improve the diagnosis effectiveness of the conventional RF classifier, the IGPR-based features should be extracted and selected before their introduction to the RF for classification.

D. FAULT DETECTION AND DIAGNOSIS USING IGPR-RF

Figure 1 shows the flowchart of the proposed FDD technique. First, the developed IGPR-RF divides the input data set (step 1) into training (used for learning) and testing (used for validation) data sets in order to distinguish between the healthy and faulty operating modes. During the training phase, the interval-valued model is firstly built using the IGPR algorithm. Second, the IGPR model extracts and selects the most effective features (step 2). Then, the RF uses the statistical IGPR parameters (selected features) for training (step 2). Finally, the classification is performed as shown in step 3. In the testing phase (step 3), the statistical IGPR parameters of the test sample data (belonging to a respective class) are extracted and selected using the IGPR technique. Then, based on the classification model computed in the training phase, the RF classifies the statistical IGPR parameters (step 3).

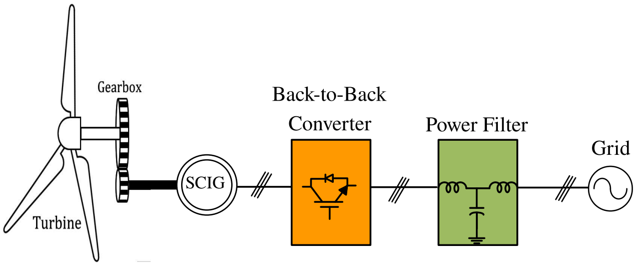

FIGURE 2. The grid-connected WEC system under study.

III. APPLICATION TO WEC SYSTEMS

The proposed FDD approach is implemented on a WEC system. Different comparative studies are investigated in this work. The proposed IGPR-RF technique is compared to IKPCA-RF, SVM, DT, NB, DA and KNN. In this work, the radial basis function (RBF) is used for all machine learning techniques with a kernel parameter σ. All the experiments are conducted using 10-fold cross-validation in the training set, after which they are applied in the testing phase. The minimum root mean-square error (RMSE) is taken as selection criterion for different machine classifiers. In the IKPCA algorithm, the parameter σ is equal to the minimum distance between the training data. The number of kernel principal components is determined using the cumulative percent variance (CPV) with a threshold equal to 95%. Naïve Bayes has an assumption that each attribute follows a normal distribution. The K value for KNN is set to 3 and for the SVM classifiers, the parameters C and σ are chosen with the lowest RMSE value and they are used for the training of SVMs for the whole data set. The parameters of IGPR model is optimized using the maximum marginal criterion. For Discriminant Analysis (DA), the regularization parameter is set as 1. For DT and RF, 50 trees are utilized. The performance is evaluated using the following criteria: Accuracy, Recall, Precision and F_1 Score [36].

A. SYSTEM DESCRIPTION

In this paper, the studied WEC system consists of a serial connection of a WT, a squirrel cage induction generator (SCIG), and a grid-connected back-to-back converter (Figure 2). The whole system is controlled to feed a fixed frequency current to the grid at unity power factor. The system parameters are presented in [23]. However, any fault in one of the above-mentioned system stages could strongly affect the power production rate [37]. As the recent studies have shown that the power electronics interface is the most sensitive WEC system stage to faults, the inverters operation should be monitored in order to ensure an effective and continuous operation. Indeed, many factors lead to the power semiconductors aging which mainly affects the time response and could lead to additional switching losses. Moreover, the excessive switching

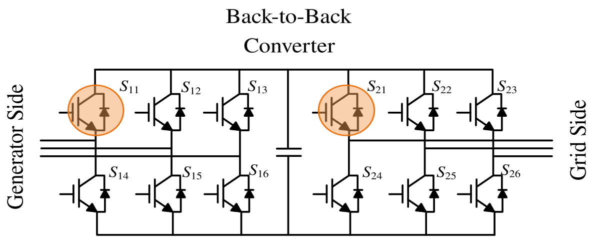

FIGURE 3. Back-to-Back converter topology.

TABLE 1. Emulated faults and their locations.

Symbol	Definition	Symbol	Definition
SC_{11}	S_{11} Short-Circuit	SC_{21}	S_{21} Short-Circuit
OC_{11}	S_{11} Open-Circuit	SC_{21}	S_{21} Open-Circuit
WO_{11}	S_{11} Wear-Out	WO_{21}	S_{21} Wear-Out

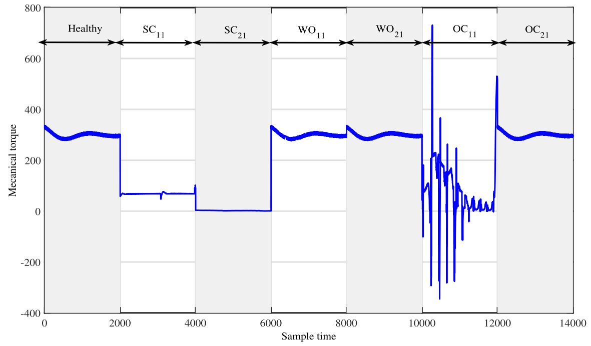

FIGURE 4. Mechanical torque under different modes.

of the transistors might be the origin of different types of faults. Therefore, it is highly recommended to early detect the transistors aging in order to prevent the overall inverter failure. For instance, the IGBT fault is preceded by an abrupt increase of the collector-emitter voltage, which is considered as a good predictive maintenance indicator [38]. In this study, the transistor aging is modeled by the increase of the internal resistance while a null value is representing the normal operating condition. In this study, the rectifier and inverter sides transistors S_{11} and S_{21} are respectively encompassed in the FDD approach (see Figure 3).

The open-circuit, wear-out, and short-circuit faults are considered in this study (Table 1). The wear-out fault is emulated by increasing the internal resistance to 2 Ω. Figures 4 to 8 show the behavior of the mechanical torque, generator speed, generator current, grid current, and DC bus voltage respectively under normal and faulty conditions.

B. DIAGNOSIS RESULTS AND COMPARISON STUDIES

Twelve variables are generated for diagnosis purposes (Table 2). Seven operating modes including one healthy and six faulty modes are used as generated simulation data series (Table 3). Each mode is adequately described over 2000 10-time-lagged samples within a $1s$ time period and 20 KHz sampling frequency [23]. The IGPR model is built by 2000 extracted samples. The IKPCA model is built under normal operating conditions using CPV criterion with 95% of confidence interval. Finally, the statistical quantities (M_{IGPR} and C_{IGPR}) obtained from the IGPR model are introduced to the RF algorithm for classification. To illustrate the classification accuracy of the developed approach, a 10-fold cross-validation scheme was adopted. The labeled data are used as inputs for all classifiers. Two types of classifiers are

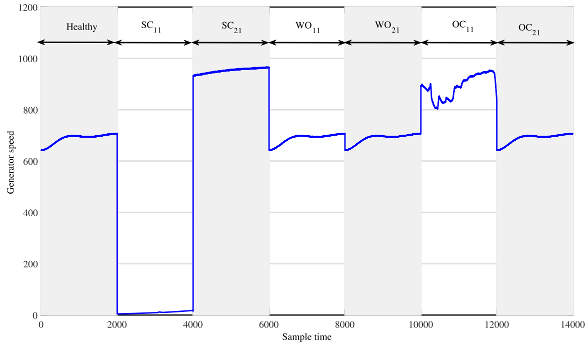

FIGURE 5. Generator speed under different modes.

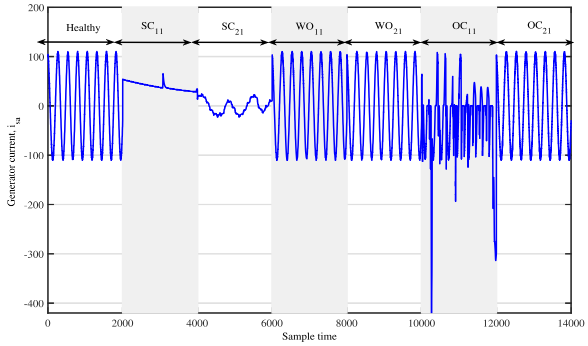

FIGURE 6. Generator current for different scenarios.

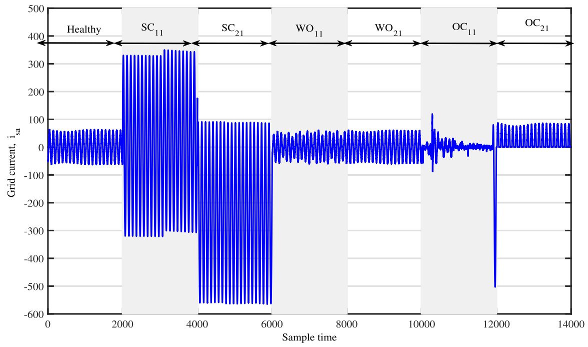

FIGURE 7. Grid current for different scenarios.

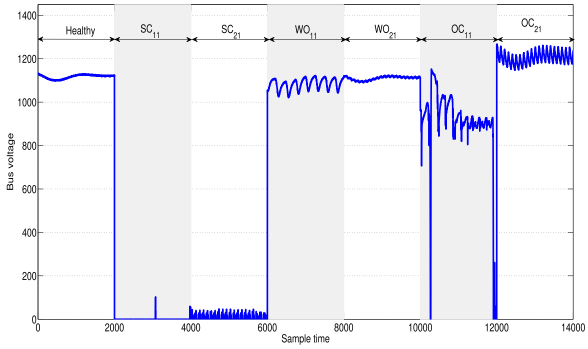

FIGURE 8. Bus voltage under different modes.

TABLE 2. Description of variables.

Variables	Descriptions
x_1	C_m: Mechanical torque (Nm)
x_2	N_g: Generator speed (tr/m)
x_3	$i_{s_{a_g}}$: Generator current phase a (A)
x_4	$i_{s_{b_g}}$: Generator current phase b (A)
x_5	I_{s_d}: Generator current along d-axis (A)
x_6	I_{s_q}: Generator current along q-axis (A)
x_7	V_{DC}: Bus voltage (V)
x_8	$P_{O_{u_t}}$: Output power (W)
x_9	$i_{s_{a_r}}$: Grid current phase a (A)
x_{10}	$i_{s_{b_r}}$: Grid current phase b (A)
x_{11}	I_{s_d}: Grid current along d-axis (A)
x_{12}	I_{s_q}: Grid current along q-axis (A)

TABLE 3. Database construction.

Class	Mode	Training	Testing
C_0	Healthy	2000	2000
C_1	SC_{11}	2000	2000
C_2	SC_{21}	2000	2000
C_3	WO_{11}	2000	2000
C_4	WO_{21}	2000	2000
C_5	OC_{11}	2000	2000
C_6	OC_{21}	2000	2000

TABLE 4. Performances comparison of different multi-class techniques.

	Methods	
Global Performance	Accuracy (Training/Testing)	F_1 score (Training/Testing)
IGPR-RF	**100 / 100**	**100 / 100**
IKPCA-RF	99.4 / 99.38	99.4 / 99.38
SVM	83.74 / 83.59	92.14/92.09
DT	73.46 / 74.14	71.81 / 73.75
NB	14.25 / 13.45	13.91 / 11.50
DA	16 / 13.11	16.13 / 10.53
KNN	77.99 / 88.30	77.75 / 88.21

TABLE 5. Confusion matrix using IKPCA-RF.

	Predicted class							
True class	C_0	C_1	C_2	C_3	C_4	C_5	C_6	Recall
C_0	1966	0	0	5	18	0	11	98.3
C_1	0	2000	0	0	0	0	0	100
C_2	0	0	2000	0	0	0	0	100
C_3	0	1	1	1996	2	0	27	99.80
C_4	13	0	1	8	1965	0	13	98.25
C_5	0	0	0	0	0	2000	0	100
C_6	4	0	3	0	7	0	1986	99.30
Precision	99.14	99.95	99.75	99.35	98.64	100	98.80	**99.38**

presented in this work: multi-class classifiers (see Table 4) and a set of one-class classifiers (Table 7). For the multi-class classifiers, Table 4 illustrates the results using the IGPR-RF, IKPCA-RF, SVM, DT, NB, DA, and KNN techniques to assess the diagnosis performance in terms of accuracy and F_1 score.

It can be noticed from Table 4 that the developed IGPR-RF technique gives a better classification accuracy compared to the IKPCA-RF and both of them outperform the raw data-based classifiers. The good performance of the developed approach is due to its effectiveness in excluding the ineffective samples and selecting the most accurate features from the predictive posterior distribution, while the IKPCA-RF uses the first principal components as inputs to the RF classifier. The SVM, DT, NB, DA and KNN classifiers are based on the direct use of the raw data.

TABLE 6. Confusion matrix using IGPR-RF.

	Predicted class							
True class	C_0	C_1	C_2	C_3	C_4	C_5	C_6	Recall
C_0	2000	0	0	0	0	0	0	100
C_1	0	2000	0	0	0	0	0	100
C_2	0	0	2000	0	0	0	0	100
C_3	0	0	0	2000	0	0	0	100
C_4	0	0	0	0	2000	0	0	100
C_5	0	0	0	0	0	2000	0	100
C_6	0	0	0	0	0	0	2000	100
Precision	100	100	100	1000	100	100	100	**100**

TABLE 7. Performances comparison of different one-class techniques.

		Method	
Class	Phase	IKPCA-RF	IGPR-RF
C_0	Training	98.33	99.99
	Testing	97.90	99.99
C_1	Training	99.99	99.99
	Testing	99.94	99.99
C_2	Training	99.99	100
	Testing	99.99	99.99
C_3	Training	98.53	99.99
	Testing	99.14	100
C_4	Training	98.42	100
	Testing	98.16	99.99
C_5	Training	99.93	100
	Testing	99.99	100
C_6	Training	98.73	99.99
	Testing	99.14	100
Average	Training	99.13	99.99
	Testing	99.18	99.99

To further assess the effectiveness of the proposed FDD method, the results are presented using the confusion matrix (Tables 5 and 6). The confusion matrix defines the number of predicted labels in columns and the number of actual labels in rows. The diagonal of the confusion matrix presents the correct classification for the seven classes (C_0 to C_6). For the testing healthy data, assigned to class C_0, the IKPCA-RF classifier (see Table 5) identifies only 1966 samples among 2000 (true positive). In addition, the detection precision is 99.14% and its recall is 98.30% which also represents the classification accauracy. So, 1.7% of misclassification is found (false alarms) for this class. A classification error of 0.200% is found for class C_3 in testing data. For the faulty case (C_4), the precision is 98.64% and the recall is 98.25% with 1.75% of misclassification for testing data set, whereas the misclassification rate for the faulty class C_6 is 0.70% and 0% for faulty classes C_1, C_2 and C_5. However, using the proposed the IGPR-RF, the precision is 100% and the recall is 100% for all cases (Table 6) which means that the classification errors are equal to 0%.

A set of one-class classifiers is presented here in order to further improve the classification capabilities of the proposed IGPR-RF strategy. For this purpose, a classifier bank that uses two classifiers based on kernel methods (IKPCA-RF, IGPR-RF) is applied to distinguish between the WEC faults. Each classifier is trained to classify a specific class with a label of 1 or −1. The classification results are presented in Table 7.

It can be seen from Table 7 that the average accuracy rate obtained using the proposed method in the training and testing cases are 99.99%. However, the IKPCA-RF indicates 99.13% of classification accuracy for the training case and 99.18% for the testing case. Thus, the developed IGPR-RF technique presents a very good accuracy in the training and testing cases compared to IKPCA-RF.

IV. CONCLUSION

In this paper, a new fault detection and diagnosis (FDD) strategy was proposed to improve the reliability of uncertain wind energy conversion (WEC) systems. To achieve a fast and reliable FDD, the most effective interval-valued features were extracted and selected using a interval GPR (IGPR) model. Then the selected interval-valued features were introduced as inputs to the RF classifier for diagnosis purposes. The simulation results were presented to prove the effectiveness of the proposed fault diagnosis strategy. The presented results showed that the IGPR-RF, with nonlinear statistical features that depend on small selected samples of the dataset, performed better than the IKPCA-RF technique that explicitly depends on the entire dataset, and way better than the conventional techniques (SVM, DT, NB, DA and KNN) using raw data. Moreover, the developed IGPR-RF technique presented a noticeable accuracy improvement compared to the IKPCA-RF where the entire dataset is used.

REFERENCES

[1] Y. Amirat, M. E. H. Benbouzid, E. Al-Ahmar, B. Bensaker, and S. Turri, "A brief status on condition monitoring and fault diagnosis in wind energy conversion systems," *Renew. Sustain. Energy Rev.*, vol. 13, no. 9, pp. 2629–2636, Dec. 2009.

[2] J. Zhang, H. Sun, Z. Sun, W. Dong, and Y. Dong, "Fault diagnosis of wind turbine power converter considering wavelet transform, feature analysis, judgment and BP neural network," *IEEE Access*, vol. 7, pp. 179799–179809, 2019.

[3] Y. Amirat, M. E. H. Benbouzid, B. Bensaker, and R. Wamkeue, "Condition monitoring and ault diagnosis in wind energy conversion systems: A review," in *Proc. IEEE Int. Electric Mach. Drives Conf.*, vol. 2, May 2007, pp. 1434–1439.

[4] D. Zhang, L. Qian, B. Mao, C. Huang, B. Huang, and Y. Si, "A data-driven design for fault detection of wind turbines using random forests and XGboost," *IEEE Access*, vol. 6, pp. 21020–21031, 2018.

[5] J.-H. Zhong, J. Zhang, J. Liang, and H. Wang, "Multi-fault rapid diagnosis for wind turbine gearbox using sparse Bayesian extreme learning machine," *IEEE Access*, vol. 7, pp. 773–781, 2019.

[6] S. Dey, P. Pisu, and B. Ayalew, "A comparative study of three fault diagnosis schemes for wind turbines," *IEEE Trans. Control Syst. Technol.*, vol. 23, no. 5, pp. 1853–1868, Sep. 2015.

[7] S. Biswas and P. K. Nayak, "A fault detection and classification scheme for unified power flow controller compensated transmission lines connecting wind farms," *IEEE Syst. J.*, early access, Jan. 23, 2020, doi: 10.1109/JSYST.2020.2964421.

[8] J. Long, J. Mou, L. Zhang, S. Zhang, and C. Li, "Attitude data-based deep hybrid learning architecture for intelligent fault diagnosis of multi-joint industrial robots," *J. Manuf. Syst.*, to be published.

[9] J. Long, S. Zhang, and C. Li, "Evolving deep echo state networks for intelligent fault diagnosis," *IEEE Trans. Ind. Informat.*, vol. 16, no. 7, pp. 4928–4937, Jul. 2020.

[10] M. Schmid, E. Gebauer, C. Hanzl, and C. Endisch, "Active model-based fault diagnosis in reconfigurable battery systems," *IEEE Trans. Power Electron.*, vol. 36, no. 3, pp. 2584–2597, Mar. 2021.

[11] M. Mansouri, M.-F. Harkat, H. N. Nounou, and M. N. Nounou, *Data-Driven Model-Based Methods for Fault Detection Diagnosis*. Amsterdam, The Netherlands: Elsevier, 2020.

[12] H. Habibi, I. Howard, and S. Simani, "Reliability improvement of wind turbine power generation using model-based fault detection and fault tolerant control: A review," *Renew. Energy*, vol. 135, pp. 877–896, May 2019.

[13] C. Du, F. Li, and C. Yang, "An improved homogeneous polynomial approach for adaptive sliding-mode control of Markov jump systems with actuator faults," *IEEE Trans. Autom. Control*, vol. 65, no. 3, pp. 955–969, Mar. 2020.

[14] L. Jiang, D. Xiang, Y. F. Tan, Y. H. Nie, H. J. Cao, Y. Z. Wei, D. Zeng, Y. H. Shen, and G. Shen, "Analysis of wind turbine Gearbox's environmental impact considering its reliability," *J. Cleaner Prod.*, vol. 180, pp. 846–857, Apr. 2018.

[15] G. Jiang, H. He, J. Yan, and P. Xie, "Multiscale convolutional neural networks for fault diagnosis of wind turbine gearbox," *IEEE Trans. Ind. Electron.*, vol. 66, no. 4, pp. 3196–3207, Apr. 2019.

[16] Y. Wang, X. Ma, and P. Qian, "Wind turbine fault detection and identification through PCA-based optimal variable selection," *IEEE Trans. Sustain. Energy*, vol. 9, no. 4, pp. 1627–1635, Oct. 2018.

[17] W. Chen, S. X. Ding, A. Haghani, A. Naik, A. Q. Khan, and S. Yin, "Observer-based FDI schemes for wind turbine benchmark," *IFAC Proc. Volumes*, vol. 44, no. 1, pp. 7073–7078, Jan. 2011.

[18] M. Kruger, S. X. Ding, A. Haghani, P. Engel, and T. Jeinsch, "A data-driven approach for sensor fault diagnosis in gearbox of wind energy conversion system," in *Proc. 10th IEEE Int. Conf. Control Autom. (ICCA)*, Jun. 2013, pp. 227–232.

[19] P. F. Odgaard and J. Stoustrup, "Unknown input observer based scheme for detecting faults in a wind turbine converter," *IFAC Proc. Volumes*, vol. 42, no. 8, pp. 161–166, 2009.

[20] A. Haghani, M. Krueger, T. Jeinsch, S. X. Ding, and P. Engel, "Data-driven multimode fault detection for wind energy conversion systems," *IFAC-PapersOnLine*, vol. 48, no. 21, pp. 633–638, 2015.

[21] Z. Hameed, Y. S. Hong, Y. M. Cho, S. H. Ahn, and C. K. Song, "Condition monitoring and fault detection of wind turbines and related algorithms: A review," *Renew. Sustain. Energy Rev.*, vol. 13, no. 1, pp. 1–39, Jan. 2009.

[22] B. Lu, Y. Li, X. Wu, and Z. Yang, "A review of recent advances in wind turbine condition monitoring and fault diagnosis," in *Proc. IEEE Power Electron. Mach. Wind Appl.*, Jun. 2009, pp. 1–7.

[23] A. Kouadri, M. Hajji, M.-F. Harkat, K. Abodayeh, M. Mansouri, H. Nounou, and M. Nounou, "Hidden Markov model based principal component analysis for intelligent fault diagnosis of wind energy converter systems," *Renew. Energy*, vol. 150, pp. 598–606, May 2020.

[24] R. Fezai, M. Mansouri, K. Abodayeh, and H. Nounou, "Online reduced Gaussian process regression based generalized likelihood ratio test for fault detection," *J. Process Control*, vol. 85, pp. 30–40, Jan. 2020.

[25] S. S. Roy, S. Dey, and S. Chatterjee, "Autocorrelation aided random forest classifier based bearing fault detection framework," *IEEE Sensors J.*, vol. 20, no. 18, pp. 10792–10800, Sep. 2020.

[26] K. Dhibi, R. Fezai, M. Mansouri, M. Trabelsi, A. Kouadri, K. Bouzara, H. Nounou, and M. Nounou, "Reduced kernel random forest technique for fault detection and classification in grid-tied PV systems," *IEEE J. Photovolt.*, vol. 10, no. 6, pp. 1864–1871, Nov. 2020.

[27] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[28] J. R. Quinlan, "Simplifying decision trees," *Int. J. Man-Mach. Stud.*, vol. 27, no. 3, pp. 221–234, Sep. 1987.

[29] L. Jiang, L. Zhang, L. Yu, and D. Wang, "Class-specific attribute weighted naive bayes," *Pattern Recognit.*, vol. 88, pp. 321–330, Apr. 2019.

[30] A. P. D. Silva and A. Stam, "Discriminant analysis," in *Reading and Understanding Multivariate Statistics*, L. G. Grimm and P. R. Yarnold, Eds. Washington, DC, USA: American Psychological Association, 1995, pp. 277–318.

[31] N. Suguna and K. Thanushkodi, "An improved k-nearest neighbor classification using genetic algorithm," *Int. J. Comput. Sci. Issues*, vol. 7, no. 2, pp. 18–21, 2010.

[32] C. Liu, L. Zhang, Y. Liao, C. Wu, and G. Peng, "Multiple sensors based prognostics with prediction interval optimization via echo state Gaussian process," *IEEE Access*, vol. 7, pp. 112397–112409, 2019.

[33] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[34] G. Biau and E. Scornet, "A random forest guided tour," *TEST*, vol. 25, no. 2, pp. 197–227, Jun. 2016.

[35] T. Lan, H. Hu, C. Jiang, G. Yang, and Z. Zhao, "A comparative study of decision tree, random forest, and convolutional neural network for spread-F identification," *Adv. Space Res.*, vol. 65, no. 8, pp. 2052–2061, Apr. 2020.

[36] K. Dhibi, R. Fezai, M. Mansouri, A. Kouadri, M.-F. Harkat, K. Bouzara, H. Nounou, and M. Nounou, "A hybrid approach for process monitoring: Improving data-driven methodologies with dataset size reduction and interval-valued representation," *IEEE Sensors J.*, vol. 20, no. 17, pp. 10228–10239, Sep. 2020.

[37] M. Rahimi, "Improvement of energy conversion efficiency and damping of wind turbine response in grid connected DFIG based wind turbines," *Int. J. Electr. Power Energy Syst.*, vol. 95, pp. 11–25, Feb. 2018.

[38] J. Due, S. Munk-Nielsen, and R. Nielsen, "Lifetime investigation of high power IGBT modules," in *Proc. 14th Eur. Conf. Power Electron. Appl.*, Aug./Sep. 2011, pp. 1–8.

MAJDI MANSOURI (Member, IEEE) received the degree in electrical engineering from SUPCOM, Tunis, Tunisia, in 2006, the M.Sc. degree in electrical engineering from ENSEIRB, Bordeaux, France, in 2008, the Ph.D. degree in electrical engineering from UTT Troyes, France, in 2011, and the H.D.R. (accreditation to supervise research) degree in electrical engineering from the University of Orleans, France, in 2019. He joined the Electrical Engineering Program, Texas A&M University at Qatar, in 2011, where he is currently an Associate Research Scientist. He is the author of more than 150 publications. He is also the author of the book *Data-Driven and Model-Based Methods for Fault Detection and Diagnosis* (Elsevier, 2020). His research interests include development of model-based, data-driven, and machine learning techniques for fault detection and diagnosis.

RADHIA FEZAI is currently an Assistant Research Scientist with the Electrical Engineering Program, Texas A&M University at Qatar. Her work focuses on the use of applied mathematics and statistics concepts to develop statistical data and model-driven techniques and algorithms for modeling, fault detection, and diagnosis with the aim of improving the operation of industrial systems.

MOHAMED TRABELSI (Senior Member, IEEE) received the B.Sc. degree in electrical engineering from INSAT, Tunisia, in 2006, and the M.Sc. degree in automated systems and the Ph.D. degree in energy systems from INSA Lyon, France, in 2006 and 2009, respectively. From October 2009 to August 2018, he was hold different Research positions with Qatar University and Texas A&M University at Qatar. Since September 2018, he has been with the Kuwait College of Science and Technology, as an Associate Professor. He has published more than 90 journal and conference papers. He is the author of two books and two book chapters. His research interests include systems control with applications arising in the contexts of power electronics, energy conversion, renewable energies integration, and smart grids.

MANSOUR HAJJI received the M.Sc. degree in electrical engineering from ESSTT, Tunis, Tunisia, in 2005, and the master's degree in electrical system and the Ph.D. degree in electrical engineering from the National Engineering School of Tunis (ENIT), Tunis, in 2008 and 2013, respectively. Since 2013, he has been with the Higher Institute of Applied Science and Technology of Kasserine, Tunisia, as an Assistant Professor. He is the author of several publications. His current research includes include electrical machine, design and control, and machine learning techniques for fault detection and diagnosis.

MOHAMED-FAOUZI HARKAT (Member, IEEE) received the engineering degree in automatic control from Badji Mokhtar–Annaba University, Annaba, Algeria, in 1996, and the Ph.D. degree from the Institut National Polytechnique de Lorraine (INPL), France, in 2003. From 2002 to 2004, he was an Assistant Professor with the School of Engineering Sciences and Technologies of Nancy (ESSTIN), France. In 2004, he joined the Department of Electronics, Badji Mokhtar–Annaba University, where he is currently a Professor. He has over 20 years of research and practical experience in systems engineering and process monitoring. He is the author of more than 100 refereed journal and conference publications and book chapters. He served as an Associate Editor on technical committees for several international journals and conferences.

HAZEM NOUNOU (Senior Member, IEEE) is currently a Professor of electrical and computer engineering with Texas A&M University at Qatar. He has more than 19 years of academic and industrial experience. He has significant experience in research on control systems, database control, system identification and estimation, fault detection, and system biology. He has been awarded several NPRP research projects in these areas. He has successfully served as the Lead PI and a PI on five QNRF projects, some of which were in collaboration with other PIs in this proposal. He has published more than 200 refereed journal and conference papers and book chapters. He has served as an Associate Editor on technical committees for several international journals and conferences.

MOHAMED N. NOUNOU (Senior Member, IEEE) is currently a Professor of chemical engineering with TAMU-Texas A&M University at Qatar. He has more than 19 years of combined academic and industrial experience. He has published more than 200 refereed journal and conference publications and book chapters. He has successfully served as the Lead PI and a PI on several QNRF projects (six NPRP projects and three UREP projects). His research interests include systems engineering and control, with emphasis on process modeling, monitoring, and estimation. He is a Senior Member of the American Institute of Chemical Engineers (AIChE).

KAIS BOUZRARA is currently a Professor of electrical engineering with the Laboratory of Automatic Signal and Image Processing, National Engineering School of Monastir, Monastir, Tunisia. He has more than 15 years of combined academic and industrial experience. He has published more than 80 refereed journal and conference publications and book chapters. His research interests include systems engineering and control, with emphasis on process modeling, monitoring, and estimation.