

Data-Driven Approach for Condition Monitoring and Improving Power Output of Photovoltaic Systems

Nebras M. Sobahi^{1,*}, Ahteshamul Haque², V S Bharath Kurukuru², Md. Mottahir Alam¹ and Asif Irshad Khan³

¹Department of Electrical and Computer Engineering (ECE), King Abdulaziz University, Jeddah 21589, Saudi Arabia

²Advance Power Electronics Research Lab, Department of Electrical Engineering, Jamia Millia Islamia, New Delhi, India

³Computer Science Department, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia

*Corresponding Author: Nebras M. Sobahi. Email: nsobahi@kau.edu.sa

Received: 08 February 2022; Accepted: 31 May 2022

Abstract: Increasing renewable energy targets globally has raised the requirement for the efficient and profitable operation of solar photovoltaic (PV) systems. In light of this requirement, this paper provides a path for evaluating the operating condition and improving the power output of the PV system in a grid integrated environment. To achieve this, different types of faults in grid-connected PV systems (GCPVs) and their impact on the energy loss associated with the electrical network are analyzed. A data-driven approach using neural networks (NNs) is proposed to achieve root cause analysis and localize the fault to the component level in the system. The localized fault condition is combined with a parallel operation of adaptive neuro-fuzzy inference units (ANFIUs) to develop a power mismatch-based control unit (PMCU) for improving the power output of the GCPV. To develop the proposed framework, a 10-kW single-phase GCPV is simulated for training the NN-based anomaly detection approach with 14 deviation signals. Further, the developed algorithm is combined with the PMCU implemented with the experimental setup of GCPV. The results identified 98.2% training accuracy and 43000 observations/sec prediction speed for the trained classifier, and improved power output with reduced voltage and current harmonics for the grid-connected PV operation.

Keywords: Condition monitoring; anomaly detection; performance evaluation; fault classification; optimization

1 Introduction

Driven by environmental and political incentives, the transition of power generation from oil-based economies to renewable energy is increasing in Middle Eastern countries. The solar photovoltaic (PV) technology, which is a long-established renewable energy source in Saudi Arabia, has encouraged achieving carbon neutrality with 50% of power generation from renewables by 2030 [1,2]. According



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

to the data analysis reported by the international renewable energy agency in 2020, the clean energy targets for 2030 aimed at 58.7 GW from renewable energy power generation where 40 GW is targeted from PV alone, 16 GW from wind, and the remaining 2.7 GW from concentrating solar power [3]. Considering the projected increase in PV power generation, an efficient means of coordination, monitoring, and maintenance approaches are required to accommodate them with the current power system infrastructure. This improves the functioning of grid-connected photovoltaic (PV) systems to ensure reliable power production and optimal energy harvesting. Further, the increasing number of data loggers and data acquisition units at various parts of the distributed generation units has inherently provided information about the operation of the system [4]. This data intrinsically handles the information about the health of the system components and provides a new approach for evaluating the operation of the power system with a great deal of economic potential for operation and maintenance. These aspects have encouraged the research on condition monitoring and power output improvement for PV systems.

Initially, the literature on condition monitoring and power output improvement is mostly developed considering the climate data at the plant location [5]. This is a cost-effective approach due to readily available metrological data but the expected energy yield has low accuracy due to the uncertain climatic conditions [6,7]. Further, the metrological data measures are combined with soft computing techniques in [8–10] to estimate energy production. Both the approaches are developed from the PV module power generation data collected using a PV system operated for normal and faulty conditions. In [11,12], the authors developed artificial intelligence and statistical data [13] based on automatic fault detection and condition monitoring of PV systems. These techniques are not optimized for localizing the faults to the component level in the PV system. In an approach proposed by Mellit [14], a standalone photovoltaic system is modelled with an adaptive neuro-fuzzy inference system to improve the power output of the PV generation to the load. In addition to the metrological data-based approaches and intelligent techniques, the simulated models based on stochastic [12], parametric [15], and non-parametric techniques [16] for the individual components of the system are developed. It is observed that these models mostly rely on the static/dynamic thresholds of the grid-connected PV system network variables to achieve condition monitoring. This resulted in drawbacks due to misclassification of the transient operating conditions as abnormalities in the system which further led to unintended trips and maintenance of the PV system. In addition to the above methods, the signal processing- techniques, which are based on the time and frequency domain were investigated in the literature to extract the fault feature from the measured converter data [17–19]. Further, the neural network-based multiple open-circuit fault diagnosis for photovoltaic inverters is developed in [20]. The research utilized the multistate data processing, and subsection fluctuation analysis approaches for practical implementation of multiple switch fault diagnosis. In [21], a temporal convolutional network-based fault diagnosis approach is developed for eliminating the problems of data noise with the conventional approaches for fault diagnosis for power converters. In general, all the statistical, data-driven, and decision-making approaches, proved to be efficient in the converter fault detection. However, their relatively slow response and inability to identify the effect of multiple faults is identified as a major drawback.

The above-discussed approaches revealed a major scientific gap between the condition monitoring and power output improvement approaches by raising questions on “how the condition monitoring approaches localize the faults/abnormalities to component level” and “how the classified condition of the PV system helps in improving the power output of the electrical network”. In light of the above observations and drawbacks, this research aims at developing a data-driven framework capable of achieving performance evaluation and improving power output for the grid-connected PV systems.

This is achieved by developing a root cause analysis approach developed with various statistical and intelligent techniques. The major aspects of the proposed framework involve:

- A data acquisition approach with a correlation approach for linear analysis between the data points to achieve efficient processing of the measured data.
- A normal behavior model through intelligent techniques for identifying the normal operation of the components.
- An anomaly detection mechanism to classify the abnormal and underperformance operation in the system and localize the faults to the component level. The anomaly data points are subjected to principal component analysis to assist in detecting the root of the anomaly.

The proposed framework is further tested with different scenarios of PV systems operated in a grid-connected environment. The remaining sections of the paper are organized as follows: Section II discusses the faults in grid-connected PV systems at the component level and their impact on the electrical network. Section III discusses the data-driven methodology for anomaly detection and performance evaluation of the grid-connected PV system. In Section IV the details of classification training for anomaly detection are discussed with measured data of various component faults in the grid-connected PV system. Section V develops the control mechanism for power output improvement with intelligent techniques and the conclusion is discussed in Section VI.

2 Failure Modes of Grid-Connected Photovoltaic Systems

Manufacturers today have addressed the technical challenges associated with the components of PV systems in grid-connected environments to achieve enhanced reliability and reduced size. But the field surveys for various grid-connected PV systems operating in different climatic zones [22,23] identified that a significant percentage of systems experience various faults and have potential performance problems [23,24]. Most of these faults are categorized in three different parts i.e., DC side, AC side, and the converter faults, for a convenient analysis [25–27]. Generally, faults on the DC side represent the PV module defects, failure of string and array interconnections, and uncertain effects of climatic conditions [28,29]. The AC side failures are registered either due to abnormalities and grid instability problems, or due to interconnection issues between the inverter, load, and grid. Further, the aging of power modules, internal failure mechanisms of converters, and control mechanism failures are categorized as the faults associated with the power electronics converters. A brief overview of commonly identified faults and their impact on the grid-connected PV systems are discussed in the further subsections.

2.1 Common Faults

Ranging from higher probability to lower probability of occurrence, the most common faults in a grid-connected PV system are identified from the literature [30] are, Inverter Internal Error (IIE), Data Acquisition System Fault (DAQF), Grid Instability (GI), and Module Fault (MF). The IIE can be observed in a PV system due to the various failure mechanisms associated with power modules such as short/open circuit failure modes, reduced dielectric strength, and increased leakage current. Generally, the fault detection system associated with the inverter can identify most of these failures with the help of the power module characteristics. This has significantly decreased the probability of inverter faults and improved its reliability. But the impact of faults on the terminal characteristics of the inverter during the aging/degradation process is difficult to identify with the integrated fault detection mechanism. Apart from the IIE, the DAQF occurs due to the failure of wired and wireless connections, and due to collateral damage to the sensors and other measurement components installed

in the system. Further, the grid instability depends on the grid quality at the point of common coupling of the PV system. This fault has a very low probability in locations with high industrial loads and a high probability in rural areas. The high probability of grid instability may result in unintentional islanding of the PV system which further leads to continuous energization of the loads causing danger for the operation and maintenance processes. The last kind of common fault is the module fault which may occur due to short/open circuits between PV strings in an array, broken cells and aging of PV modules, and due to other external conditions like periodic shading. The short circuit faults happen due to the involvement of a short circuit between the PV strings in an array and the ground, whereas the broken cell and aging of modules are related to the accidental impact of objects and hotspots on the PV modules, and the average annual degradation of the PV module material respectively [31]. Further, the periodic shading is not a fault but it impacts the power generation capability of the PV module or array by shadows which can deteriorate the surface area of modules. Therefore, from the identified common faults, their effect in terms of energy loss in the grid-connected PV system is summarized in Fig. 1.

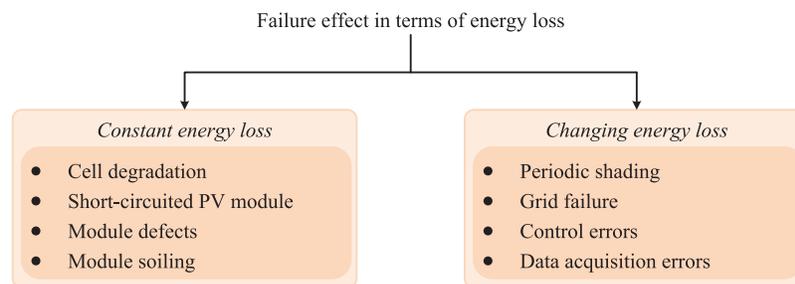


Figure 1: Effect of common faults in terms of energy loss on grid-connected photovoltaic system

2.2 Impact of PV System Faults on Electrical Network

The grid integration of PV systems is usually achieved to improve the performance of the electric network through grid feeding and grid supporting modes. This helped in improving the voltage profile and reducing the power loss in the network. However, the uncertain operation of PV systems due to varying metrological data, aging/degradation of components, and other external factors has imposed several negative impacts on the electrical network. These negative impacts resulted in voltage fluctuation problems, power mismatch, harmonics, overloading of feeder, and malfunctioning of relays and protection devices. The literature corresponding to the negative impacts of grid-connected PV system failures on the electrical network has classified the problem into three main categories as discussed in Fig. 2.

3 Data-Driven Methodology

The failure modes of PV systems and their impact on the electrical network discussed in section II identified the need for the development of a condition monitoring and maintenance management approach. Considering this, a root cause analysis approach (RCAA) is proposed to improve the operation and maintenance activities of the grid-connected PV system. The proposed approach works by analyzing the current and past operating conditions of the PV systems and has advantages in localizing the abnormality in the system to the component level. This reduces the maintenance downtime and increases the system availability time for operation and hence, bringing out better efficiency and higher savings.

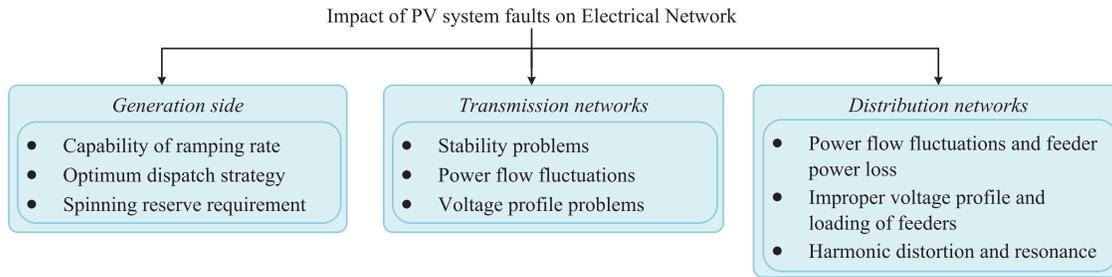


Figure 2: Negative impacts of grid-connected photovoltaic system faults on the electrical network

The RCAA is developed on a generic data-driven framework for anomaly detection and performance analysis as shown in Fig. 3. Initially, the data of abnormality conditions for different components of the PV system connected to the electrical network are identified from historical failure moments and simulated conditions. Further, a normal behavior model is derived through the data of all the components by using a data-driven model. Furthermore, the abnormal behaviors of the systems currently operating the electrical network are tested with the normal behavior model and the RCAA is performed on the identified condition of the system.

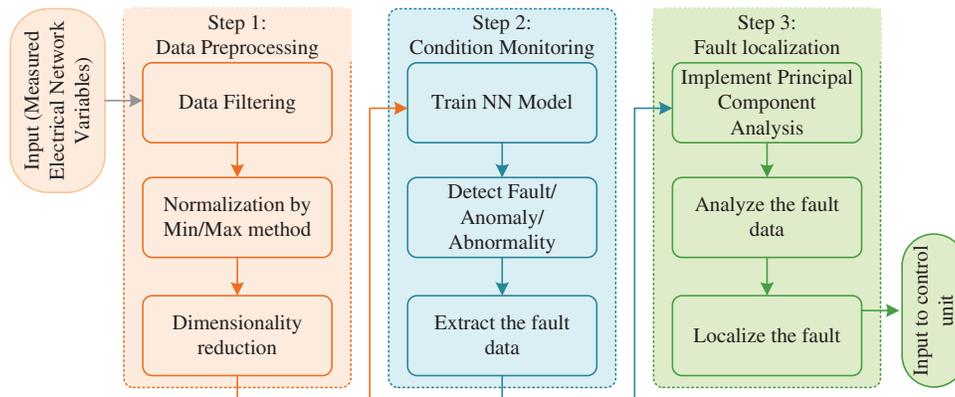


Figure 3: Generic data-driven anomaly detection and performance analysis framework

Step1: Pre-processing: The raw data recorded with the sensors installed on components of the grid-connected PV systems provide a large set of information and noise. These large data sets can be complex to train and may result in time-consuming and inefficient anomaly detection approaches. Hence, to enhance the computational efficiency of the RCCA, the data needs to be processed for achieving anomaly detection and performance analysis. For these reasons, different statistical, and cause and effect methods are discussed in the literature to improve the analysis. These approaches are proved to have higher efficiencies when applied with complex signals and also provide data protection (no loss of information) [32,33] which will be an added advantage especially while dealing with the aspects of condition monitoring. Further, it is identified that the statistical approaches have drawbacks with physical knowledge of the system and hence, the cause-and-effect relationship methods like Pearson correlation are used to improve the analysis. Before processing the raw data with the Pearson

correlation approach, it is scaled to provide a positive normalization using (1) as

$$x_i = \frac{\varphi [x_i - \text{Min}(x)]}{\text{Max}(x) - \text{Min}(x)} \quad (1)$$

where $x = \{x_1, x_2, \dots, x_n\}$ corresponds to the set of variables, φ identifies the maximum limit for the range of normalization, and n corresponds to the total measurement data points.

Further, the Pearson correlation provides a linear relationship between two variables to present a piece of elementary information [34] which is advantageous at the preliminary stage of the analysis. This information corresponds to the sign of the coefficient which signifies the linearly positive and negative output relationship among the variables, and the magnitude of the coefficient which signifies the small, medium, and large mean correlation level between the variables. The expression for achieving linear relation between variables using Pearson correlation is given as

$$pc_{X,Y} = \frac{\sum_{i=1}^o (x_i - x') (y_i - y')}{\sqrt{\sum_{i=1}^o (x_i - x')^2} \sqrt{\sum_{i=1}^o (y_i - y')^2}} \quad (2)$$

where $X = \{x_1, x_2, \dots, x_n\}$ and $Y = \{y_1, y_2, \dots, y_n\}$ are two comparable vectors, x_0 and y_0 are the mean values of their vectors X and Y respectively, and o indicates the total observations.

Step2: Anomaly Detection: To analyze the component performance in the system, a normal behavior model is simulated, and its output is compared with the monitored behavior. This aids in differentiating and detecting the anomaly in the system operation. The generalized approach for the normal behavior model with the multilayer neural network perceptron approach is shown in Fig. 4.

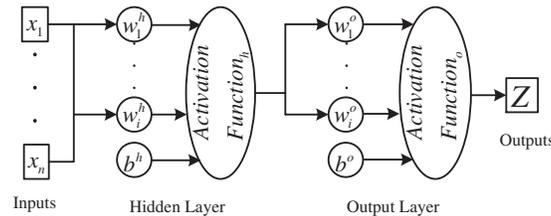


Figure 4: Structure of a feed-forward multilayer perceptron neural network

The neural network (NN) approach provides great flexibility while handling nonlinear complex systems and developing the normal behavior model for the component. The real-time measurements from the sensors at the grid-connected PV system components are the input variables of the NN and the target is labeled based on the objective of the model. The optimum configuration for developing the NN model is achieved by testing the various possible combinations and by using the pruning algorithm. Further, the optimum configuration is trained with the scaled conjugate gradient approach to tuning the biases and weights for minimizing the objective function of the NN. Further, the activation functions in the network structure are estimated as

$$\text{Activation Function}_h = f_1 (w_i^h x_i + b^h) \quad (3)$$

$$\text{Activation Function}_o = f_2 (w_i^o v_i^h + b^o) \quad (4)$$

where w_i^h and w_i^o are the weights of neurons in the hidden and output layer respectively, b^h and b^o are the bias values of hidden and output layers respectively, and v_i^h identifies the relationship between the hidden layer output and the output layer. Further, the sigmoid function with hyperbolic tangent form calculates the activation function (f_1) at the outputs of the hidden layer as [35].

$$f_1 (w_i^h x_i + b^h) = \frac{2}{1 + e^{-2 (w_i^h x_i + b^h)}} - 1 \quad (5)$$

Similarly, a linear function calculates the activation function (f_2) to estimate the outputs at the output layer as

$$f_2 (w_i^o v_i^h + b^o) = w_i^o v_i^h + b^o \quad (6)$$

The normal behavior model developed with the trained NN structure provides a deviation signal (D_s) as

$$DS = |PredictedValue - MeasuredValue| \quad (7)$$

The D_s represent anomalies in the grid-connected system that are developed from the model inaccuracy, component wear-out failures, transient abnormal behaviors, and other random failures. For a positive value of the D_s , the underperformance of the system can be identified, and for a negative value, the abnormality information can be obtained. Further, to minimize the misclassifications, and false predictions, a filtration approach [36] is adapted. In this approach, a dynamic confidence band is set for the consecutive data points such that any datapoint should not exceed four times the average of its predecessor and successor. Furthermore, based on the deviation identified, the anomaly is detected, and the data is sent for root cause analysis.

Step3: Root Cause Analysis: To deal with the correlated data and achieve root cause analysis, the principal component analysis (PCA) approach is adopted. This approach defines a new system of coordinates to transform the correlated data into uncorrelated linear data. The extracted data points from the anomaly detection are provided as input to the PCA. Initially, the PCA approximates a sufficient number of input variables to model the system without the loss of any information. This helps in improving the accuracy of analysis as the resultant variables are aimed at capturing the variability, and the highest level of change in the data. Further, PCA estimates the covariance matrix $[C]_{n \times n}$ of the data as

$$[C]_{n \times n} = [cov(x, y)] \quad (8)$$

$$cov(x, y) = \frac{1}{n-1} \sum_{i=1}^n (X_i - X') (Y_i - Y') \quad (9)$$

From the covariance matrix, the eigenvectors and eigenvalues are extracted, and the eigenvector with the highest eigenvalue is defined as a principal component. Besides, the first principal component of the PCA handles the highest variability in the data [12,34]. Further, all the principal components are combined to create a matrix of vectors known as feature matrices. To get the outcome for the PCA, the transpose of original data is multiplied by the feature matrix, and the resultant new principal components are obtained. As an outcome of the PCA, the principal component coefficients, the variance of the newly defined principal components and their percentage, and the original data in the mapped coordinate system are obtained.

4 Photovoltaic System Anomaly Detection

To assess the operation of the proposed framework with the grid-connected PV systems, a single-phase grid-connected PV system is simulated with the MATLAB/Simulink environment. To model the system, the inputs to the PV array i.e., the irradiance and the temperature are considered to be constant. A degradation factor of 4% is applied to the PV output to account for the aging effect of PV modules; therefore, a gain of 0.96 is applied to the current of the PV array. The converter consists of a single-phase full-bridge inverter operating with an outer voltage loop and inner current loop-based pulse width modulation control (PWM) control strategy. The input at the DC link of the inverter is 400V DC and the output voltage is 230V AC. Further, an LCL filter is used at the inverter output to smoothen out the harmonics caused during the switching process in the inverter. All the monitoring and measurement units are placed at inverter terminals, and the point of common coupling of the system to analyze the impact of four different failure mechanisms i.e., the module faults, converter faults, grid abnormality, and control unit failure in the system. To generate the failure or short-circuiting of PV modules, the number of PV modules in a series of Simulink PV Component series-connected modules per string is changed during the running of the simulation to reduce the string size. This value gets normal once the fault is removed. Similarly, to enable or disable single or multiple strings, a constant value of one or zero is multiplied by the input irradiance and temperature to enable or disable the string. The DC link voltage control unit active or inactive is done in the same way, a constant one or zero is given input to a selector block which enables or disables the output of the control unit. Further, the partial shading or soiling loss means less irradiance is being incident on the surface of the PV module; therefore, the gain is multiplied by the irradiance on the string which reduces the irradiance representing partial shading or soiling effect on that particular string. To simulate the converter faults, the power modules are open-circuited in one leg at a time and the corresponding characteristics are analyzed. The grid side abnormalities are modeled by introducing the low voltage and low-frequency faults at the point of common coupling which further impact the control operation leading to synchronization failure in the system. The simulation schematic of the grid-connected PV system is shown in Fig. 5 and the simulation parameters are presented in Tab. 1.

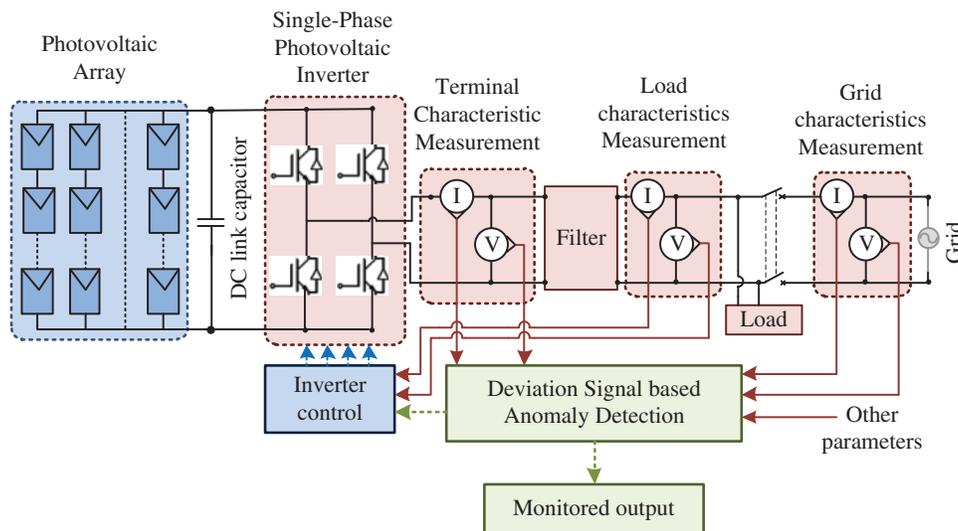


Figure 5: Schematic of single-phase grid-connected photovoltaic system with anomaly detection framework

Table 1: Specifications of simulated single-phase grid-connected photovoltaic system

System parameter		Specification
Photovoltaic array		10 kW
Photovoltaic module characteristics	Power	250 Wp
	Open circuit voltage	37.32 V
	Maximum powerpoint voltage	30.84 V
	Short circuit current	8.84 A
	Maximum powerpoint current	8.11 A
DC link capacitor		4000 μ f
Inverter rating		10 kW/12.5 kVA (Transformerless topology)

Initially, four different faults and one normal behavior model of the grid-connected PV system are simulated to obtain the data required for training the NN classifier. To simplify the classification process, the rate of change of inverter terminal voltage $\left(\frac{dv_i}{d}\right)$ and current $\left(\frac{di_i}{dt}\right)$, rate of change of load voltage $\left(\frac{dv_l}{dt}\right)$ and current $\left(\frac{di_l}{dt}\right)$, rate of change of grid voltage $\left(\frac{dv_g}{dt}\right)$ and current $\left(\frac{di_g}{dt}\right)$, rate of change of system frequency $\left(\frac{df}{dt}\right)$, zero crossing for the load voltage and grid voltages, maximum and minimum peaks of load and grid voltages, and standard deviation of the rate of change of system frequency (σ_f) are measured as the deviation signals for a fixed window length (ΔT) for all the faults and normal behavior. The standard deviation of the rate of change in system frequency (σ_f) is calculated as

$$\sigma_f = std \left\{ \frac{df(\tau)}{dt}; \tau \in [t - \Delta T, t] \right\} \quad (10)$$

The identified deviation signals are trained with the NN classifier by adapting the structure shown in Fig. 4. The network structure consists of 14 inputs, one hidden layer, and one output layer. The output layer is targeted at classifying five different classes i.e., Module faults, Converter Faults, Grid Abnormality, and Control unit failure. All the deviation signal data and the target labels correspond to form a feature matrix of size 8000×14 , where each target label has an assigned data of size 1600×14 . Further, each layer has a different activation function as discussed in Section III. To improve the sensitivity of the deviation signals to the input parameter changes of the neural network, a confidence interval of 95% is applied. This interval intensifies the impact of changes in the signal and simplifies it for anomaly detection model preparation. To assess the impact of this process, classifier training results for data without confidence interval and with confidence interval are shown in Figs. 6 and 7 respectively. From the above figures, it is identified that the data without confidence interval limits (Figs. 6a–6c) has higher misclassification samples than the data with confidence intervals (Figs. 7a–7c). This is because the data without confidence intervals have many small-valued deviation signal data points which are accumulated in the feature space during the training process and reduce the accuracy of the model. This also increases the difficulty in identifying the input data pattern, especially

while testing conditions. Further, the results of the classifier training for data without confidence intervals were 85.2% training accuracy, with 1183 misclassification costs. The green diagonal in Fig. 6a, identifies truly classified samples i.e., 6817 samples, and the upper and lower diagonal elements correspond to the misclassified samples (misclassification cost). Further, based on the trained and misclassified samples, the true positive rate, and false-negative rate are estimated as shown in Fig. 6b. Further, the positive predictive value and the false detection rate are estimated as shown in Fig. 6c to identify the accuracy of the trained classifier. The classifier training with no confidence interval data set has a training time of 23.276 s, and a prediction speed of 2300 observations per sec.

True Class	Fault1	939	636			25
	Fault2		1600			
	Fault3		231	1269	100	
	Fault4			105	1495	
	Normal	86				1514
		Fault1	Fault2	Fault3	Fault4	Normal
		Predicted Class				

(a) Truly and falsely classified samples for deviation signal data set without confidence interval bounds

True Class	Fault1	58.7%	39.8%			1.6%	58.7%	41.3%
	Fault2		100.0%				100.0%	
	Fault3		14.4%	79.3%	6.3%		79.3%	20.7%
	Fault4			6.6%	93.4%		93.4%	6.6%
	Normal	5.4%				94.6%	94.6%	5.4%
		Fault1	Fault2	Fault3	Fault4	Normal	TPR	FNR
		Predicted Class						

(b) True positive rate and false -negative rate for deviation signal data set without confidence interval bounds

True Class	Fault1	91.6%	25.8%			1.6%
	Fault2		64.9%			
	Fault3		9.4%	92.4%	6.3%	
	Fault4			7.6%	93.7%	
	Normal	8.4%				98.4%
		Fault1	Fault2	Fault3	Fault4	Normal
		Predicted Class				
	PPV	91.6%	64.9%	92.4%	93.7%	98.4%
	FDR	8.4%	35.1%	7.6%	6.3%	1.6%

(c) Positive prediction value and false detection rate for deviation signal data set without confidence interval bounds

Figure 6: Feed-forward machine learning classifier training for deviation signal data set without confidence interval bounds

True Class	Fault1	1600				
	Fault2	43	1529	28		
	Fault3		10	1574	16	
	Fault4			5	1595	
	Normal	5	34			1561
		Fault1	Fault2	Fault3	Fault4	Normal
		Predicted Class				

(a) Truly and falsely classified samples for deviation signal data set with 95% confidence interval bounds

True Class	Fault1	100.0%					100.0%	
	Fault2	2.7%	95.6%	1.8%			95.6%	4.4%
	Fault3		0.6%	98.4%	1.0%		98.4%	1.6%
	Fault4			0.3%	99.7%		99.7%	0.3%
	Normal	0.3%	2.1%			97.6%	97.6%	2.4%
		Fault1	Fault2	Fault3	Fault4	Normal	TPR	FNR
		Predicted Class						

(b) True positive rate and false-negative rate for deviation signal data set with 95% confidence interval bounds

True Class	Fault1	97.1%				
	Fault2	2.6%	97.2%	1.7%		
	Fault3		0.6%	97.9%	1.0%	
	Fault4			0.3%	99.0%	
	Normal	0.3%	2.2%			100.0%
		Fault1	Fault2	Fault3	Fault4	Normal
PPV	97.1%	97.2%	97.9%	99.0%	100.0%	
FDR	2.9%	2.8%	2.1%	1.0%		
		Predicted Class				

(c) Positive prediction value and false detection rate for deviation signal data set with 95% confidence interval bounds

Figure 7: Feed-forward machine learning classifier training for deviation signal data set with 95% confidence interval bounds

Similarly, the results are shown in Figs. 7a–7c identify the training accuracy of the classifier with a confidence interval trimmed data set. The trained classifier has an accuracy of 98.2% with a total misclassification cost of 141 samples. The truly classified samples in Fig. 7a are 7859 samples, and the falsely classified samples are 141 samples. Further, the true positive rate, false-negative rate, and positive predictive value, false detection rate are estimated as shown in Figs. 7b and 7c for analyzing the classifier training performance. The trained classifier with a confidence interval trimmed data set has a training time of 12.86 s and a prediction speed of 43000 observations per sec. To assess the superiority of the developed fault classification approach, a brief comparison between the different fault detection techniques in the literature is performed as shown in Tab. 2.

Table 2: Parametric comparison of different fault diagnosis approaches

Reference	Application	Fault type	Parameter utilized	Algorithm	Detection time	Accuracy
[17]	Power inverters	Open-circuit fault	Phasor representation of inverter current	Time-domain analysis	8.33 <i>ms</i>	–
[20]	Photovoltaic inverter	Multiple open-circuit faults	Output current feature analysis	Artificial neural network	–	–
[21]	Power converters	Open-circuit faults	Phase A output voltage	Temporal convolutional network	–	100%
Conventional approaches	Power inverters	Switch open-circuit, short-circuit, filter failure, and sensor failure	phase current and error residue	Extreme learning machine with single hidden layer feedforward neural network	15 s (Training and testing)	91.3%
	Power inverters	Switch open-circuit, short-circuit, filter failure, and sensor failure	phase current and error residue	Random forest (Boosted trees)	0.2 <i>ms</i>	79.4%
	Power inverters	Switch open-circuit, short-circuit, filter failure, and sensor failure	phase current and error residue	k-Nearest neighbors	0.41 <i>ms</i>	83.7%
	Power inverters	Switch open-circuit, short-circuit, filter failure, and sensor failure	phase current and error residue	Deep deterministic policy gradient	0.58 <i>ms</i>	91.4%

(Continued)

Table 2: Continued

Reference	Application	Fault type	Parameter utilized	Algorithm	Detection time	Accuracy
Proposed approach	Photovoltaic system	Module faults, Converter faults, Grid abnormality, and control unit failure	Voltage, current, and frequency	feed-forward multilayer perceptron neural network	0.21 ms	98.2%

Further, the trained classifier is connected in a feedback loop with the power management and inverter control unit to stabilize the system operation and improve the power output of the grid-connected PV system.

5 Optimized Control Unit

The trained classifier in the previous section can be used to identify the operating state of the grid-connected PV system. The identified operating state can be further provided as an input to the inverter control unit to achieve improved operation of the system. In this research, the classified state information is combined with the power mismatch-based control unit (PMCU) to achieve improved power output for the system. The PMCU is developed with two adaptive neuro-fuzzy inference units (ANFIUs) [37,38] by providing the difference between the measured and rated active and reactive power, and the rate of change of difference, along with the classified state of the system as an input. The internal operation of the ANFIU is similar to the structure and operation of NN but with two additional layers. These additional layers handle the information related to the fuzzy controller i.e., membership grade of linguistic variables, and firing strength of the rules. Further, the output of the ANFIU identifies the reference values for active and reactive current to improve the power output of the system [39,40]. The structure of the PMCU to achieve power output improvement with the monitored system operating condition is shown in Fig. 8, and the parameters for the operation of the ANFIU are discussed in Tab. 3.

Further, the trained classifier and the developed control unit are tested with a single-phase grid-connected laboratory setup as shown in Fig. 9. To realize the operation of the PV modules operating in an array of 10 kW, the Agilent Keysight solar simulator is connected to the network. The interconnection between the PV simulator and the grid is achieved with a Semikron bridge operating as an H-bridge inverter. The proposed condition monitoring framework is implemented with the Altera Cyclone IV field-programmable gate array and the control of the grid-connected inverter along with the PMCU is developed with the Typhoon Hardware-in-Loop. The voltage, current, and active/reactive during normal operation and a constant energy loss fault type are tested on the laboratory setup. To assess the superiority of the proposed monitoring and power output improvement framework, the results obtained are compared with the results from the action of the conventional inverter control unit. The results of the conventional inverter control unit are shown in Fig. 10. From Fig. 10a, the action of the controller for normal operation of the system operating with a 10 kWp resistive load is observed. Further, in Fig. 10b, the action of a constant energy loss fault is tested by

implementing a short-circuit fault with the simulated PV array at 190 ms. Here, the transients in voltage and current can be identified and the active and reactive power varies for a high transient resulting in the increase of the active current component for around 15 ms. In this case, the reactive power decreases momentarily for 15 ms and then starts to improve the voltage profile.

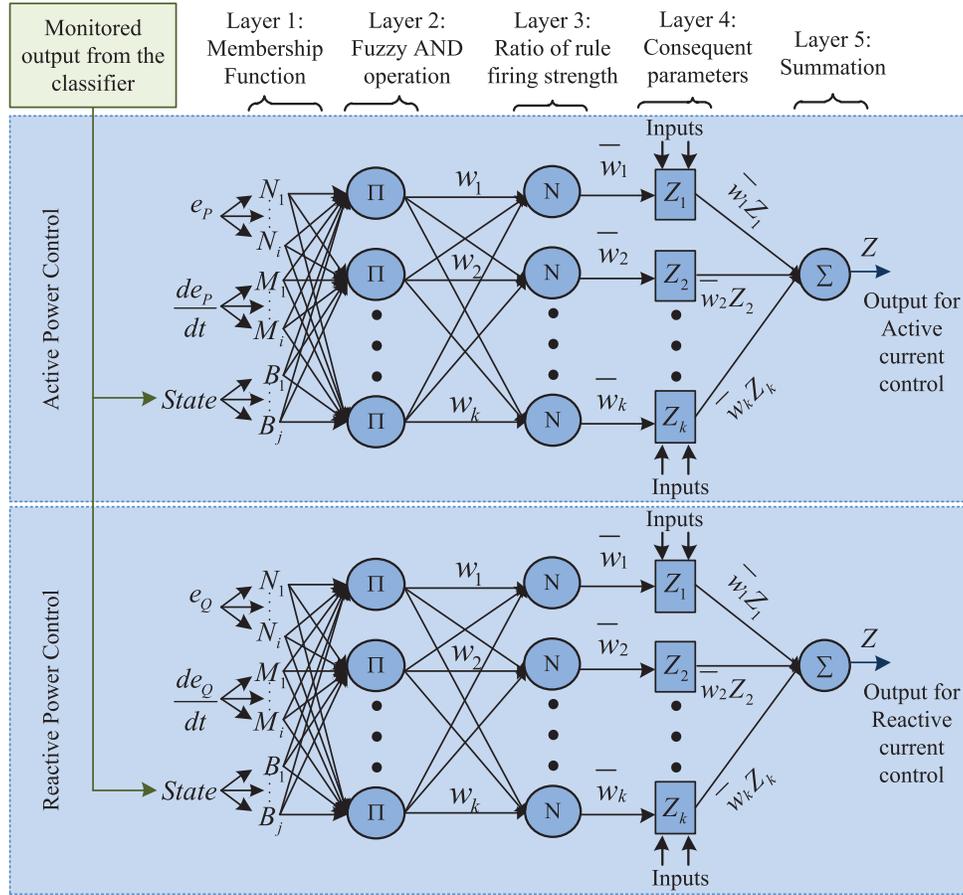


Figure 8: Adaptive fuzzy inference unit network structure for power mismatch-based control unit

Table 3: Adaptive fuzzy inference unit parameters for power mismatch-based control unit

Control parameter	Value
Number of inputs for each unit	3
Membership function (MF)	Gaussian
Number of MFs	5
Algorithm	Hybrid learning algorithm
Number of iterations	100 to 150
Fuzzy system	Sugeno type first-order system

(Continued)

Table 3: Continued

Control parameter	Value
Output type	Linear
Step size	Initial 1.1, decreased to 0.9
Nodes	55
Fuzzy rules	15
Data/parameter ration	≈ 84

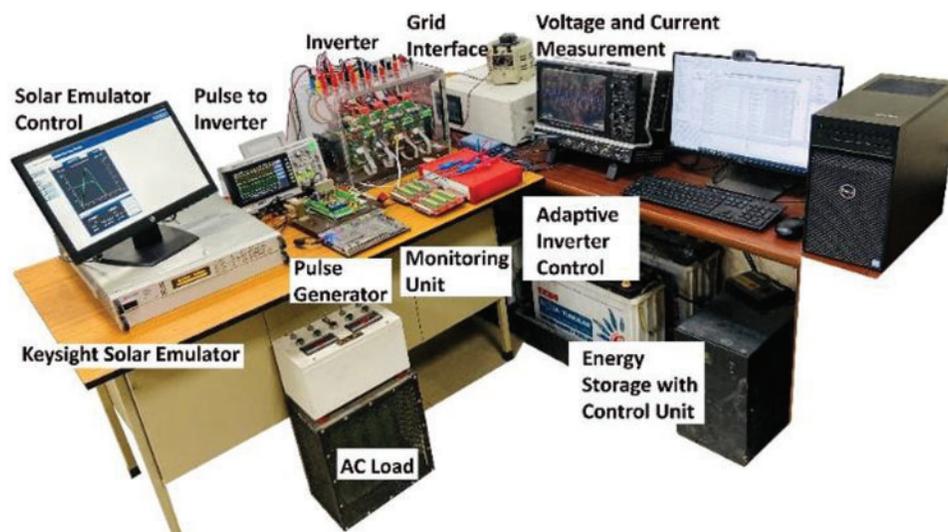


Figure 9: Experimental setup of single-phase grid-connected system with anomaly detection and power mismatch-based control unit

From the result, it is identified that the action of the conventional control approach tries to improve the voltage amplitude but has no effect on the harmonics present in both the voltage and current. This heavily affects the power out of the system. A similar condition is tested in Fig. 10c where, the constant energy loss fault is cleared, and the controller operates with the system in a grid feeding mode. Here, it is observed that due to a sudden change in the condition of the system, the controller fails to accommodate the system operation with the restoration process and generates multiple transients. This operation has a severe impact on the active and reactive power output of the system and may result in the unstable operation of the system. Further, the Fast Fourier Transform spectrum of the voltage and current during the constant energy loss fault is shown in Fig. 10d. The results identified high harmonics with an average value of around 20 dB and 60 dB for voltage and current respectively throughout the fixed window length of 500 ms.

Further, the action of the developed framework with the PMCU connected in the feedback loop with the conventional inverter control is shown through the results in Fig. 11. The results in Fig. 11a identify the normal operation of the system operating with a 10 kWp resistive load. In Fig. 11b, the action of a constant energy loss fault is tested by implementing a short-circuit fault with the simulated PV array at 140 ms. Here, the detection unit identifies the fault and activates the PMCU to restore the voltage magnitude by achieving smooth reactive power control in the system. The monitored action of

the controller also eliminates the transients in voltage and current which heavily improves the power out of the system. Further, a similar condition is tested in Fig. 11c where, the constant energy loss fault is cleared, and the system restores the normal operation. Here, again with the timely action of the classifier in identifying the system operating state, a smooth restoration of active power with fewer harmonics and voltage and current is achieved. Further, the Fast Fourier Transform spectrum of the voltage and current during the constant energy loss fault is shown in Fig. 11d. The results identified an average value of around 16 dB and 40 dB harmonic magnitude for voltage and current respectively throughout the fixed window length of 500 ms. From the results, it is observed that the developed framework handles the faults and normal operation of the system efficiently and improves the power output.

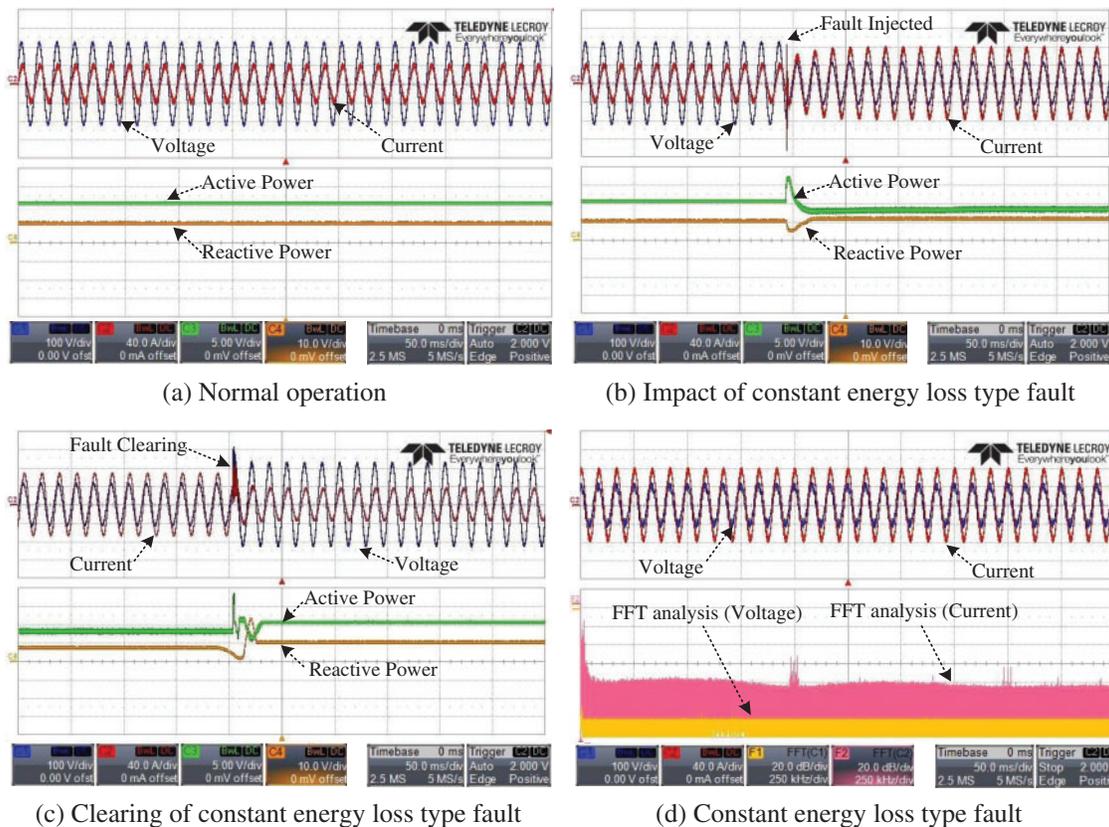


Figure 10: Electrical network variables of single-phase grid-connected photovoltaic system operating with the conventional inverter control approach. (a) Normal operation (b) impact of constant energy loss type fault (c) clearing of constant energy loss type fault and (d) constant energy loss type fault

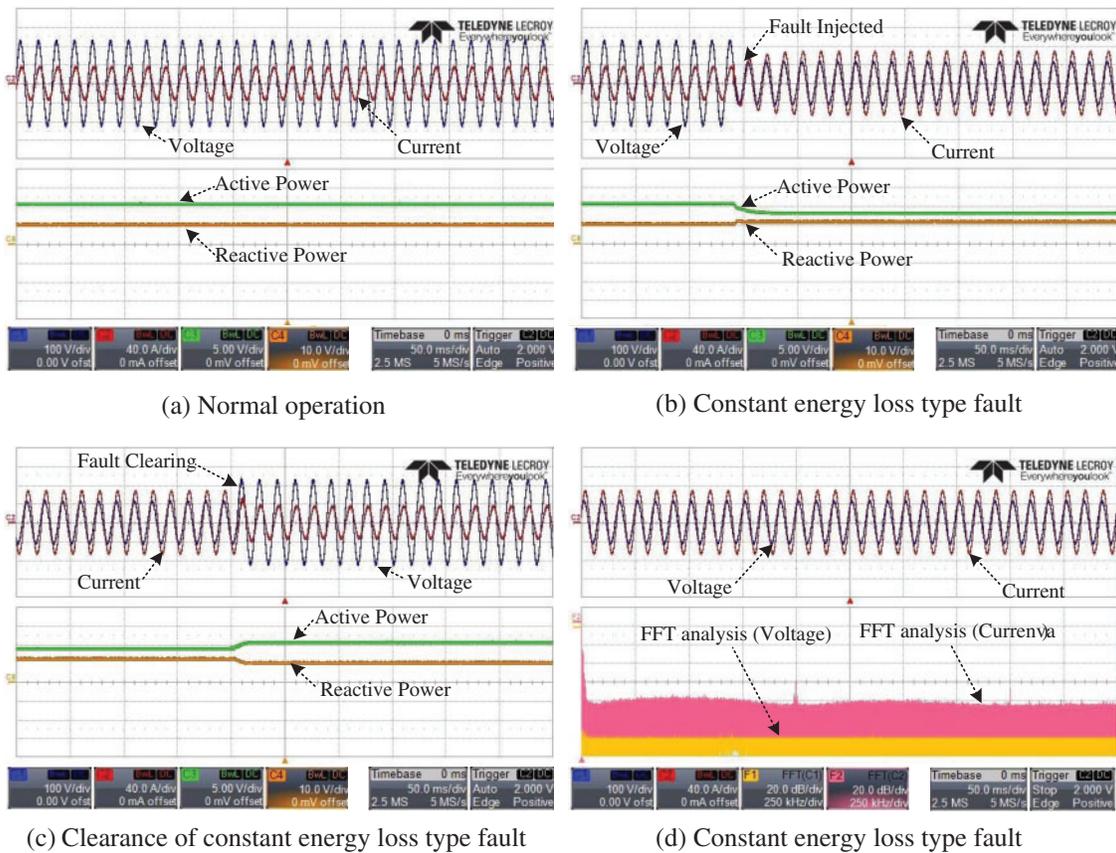


Figure 11: Electrical network variables of single-phase grid-connected photovoltaic system operating with the anomaly detection and power mismatch-based control unit framework. (a) Normal operation (b) constant energy loss type fault (c) clearance of constant energy loss type fault and (d) constant energy loss type fault

6 Conclusion

This paper establishes an approach for identifying faults at the component level in a grid-connected PV and utilizing the information for improving the operation and power output of the system. Initially, all the possible common faults, abnormalities, and degradation aspects associated with the components in a grid-connected PV system are categorized to differentiate their impact on normal behavior. Further, a 10 kWp single-phase grid-connected PV system is simulated in MATLAB/Simulink, and the deviation signals associated with the abnormalities are identified and the linear correlation between the data point is established to pre-process the data. To achieve efficient training and classification, the deviation signal data is subjected to principal component analysis with a 95% confidence interval. This data is further trained with the feed-forward multilayer perceptron neural network classifier and the training accuracy is observed to be 98.2%. To assess the superiority of the developed framework the training of the abnormality detection unit is simultaneously developed with deviation signal data with confidence interval limits and the accuracy is identified to be 85.2%. Further, the trained classifier output is combined with a power mismatch-based control unit which is developed with two parallel operated adaptive neuro-fuzzy inference units. The developed frameworks

improve the voltage and current profile, especially during the faults, and improve the power outputs of the system.

Acknowledgement: The authors extend their appreciation to the Deputyship for Research & Innovation, Ministry of Education in Saudi Arabia for funding this research work through the project number “IFPHI-021–135–2020” and King Abdulaziz University, DSR, Jeddah, Saudi Arabia.

Funding Statement: Funding for this study was received from the Deputyship for Research & Innovation, Ministry of Education in Saudi Arabia through the project number “IFPHI-021–135–2020” and King Abdulaziz University, DSR, Jeddah, Saudi Arabia.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] E. M. Barhoumi, P. C. Okonkwo, M. Zghaibeh, I. Ben Belgacem, T. Abdullah Alkanhal *et al.*, “Renewable energy resources and workforce case study Saudi Arabia: Review and recommendations,” *J. Therm. Anal. Calorim.*, vol. 141, no. 1, pp. 221–230, 2020.
- [2] L. Collins, “We will be pioneering’: Saudi Arabia reveals 50% renewables goal by 2030, but is that realistic? | Recharge,” Recharge. <https://www.rechargenews.com/energy-transition/we-will-be-pioneering-saudi-arabia-reveals-50-renewables-goal-by-2030-but-is-that-realistic-/2-1-954094> (accessed Mar. 06, 2021).
- [3] IRENA (2020), “Power sector planning in arab countries: Incorporating variable renewables,” Abu Dhabi, 2019.
- [4] T. Yang, “ICT technologies standards and protocols for active distribution network,” in *Smart Power Distribution Systems*, vol. 1, Academic Press, Elsevier, pp. 205–230, 2019.
- [5] C. Ventura and G. M. Tina, “Development of models for on-line diagnostic and energy assessment analysis of PV power plants: The study case of 1 MW sicilian PV plant,” *Energy Procedia*, vol. 83, pp. 248–257, 2015.
- [6] M. Muselli, G. Notton, J. L. Canaletti and A. Louche, “Utilization of meteosat satellite-derived radiation data for integration of autonomous photovoltaic solar energy systems in remote areas,” *Energy Conversion and Management*, vol. 39, no. 1–2, pp. 1–19, 1998.
- [7] F. Tahri, A. Tahri and T. Oozeki, “Performance evaluation of grid-connected photovoltaic systems based on two photovoltaic module technologies under tropical climate conditions,” *Energy Conversion and Management*, vol. 165, pp. 244–252, 2018.
- [8] M. Shahzad, K. V. S. Bharath., M. A. Khan and A. Haque, “Review on reliability of power electronic components in photovoltaic inverters,” in *2019 Int. Conf. on Power Electronics, Control and Automation (ICPECA)*, New Delhi, India, pp. 1–6, 2019.
- [9] H. ur Rehman, T. Korvola, R. Abdurafikov, T. Laakko, A. Hasan *et al.*, “Data analysis of a monitored building using machine learning and optimization of integrated photovoltaic panel, battery and electric vehicles in a central european climatic condition,” *Energy Conversion and Management*, vol. 221, pp. 113206, 2020.
- [10] V. S. B. Kurukuru, A. Haque, A. K. Tripathi and M. A. Khan, “Condition monitoring of IGBT modules using online TSEPs and data-driven approach,” *International Transactions on Electrical Energy Systems*, vol. 31, no. 8, pp. 1–23, 2021.
- [11] R. Fazai, K. Abodayeh, M. Mansouri, M. Trabelsi, H. Nounou *et al.*, “Machine learning-based statistical testing hypothesis for fault detection in photovoltaic systems,” *Solar Energy*, vol. 190, pp. 405–413, 2019.
- [12] B. Taghezouit, F. Harrou, Y. Sun, A. H. Arab and C. Larbes, “Multivariate statistical monitoring of photovoltaic plant operation,” *Energy Conversion and Management*, vol. 205, pp. 112317, 2020.
- [13] F. Harrou , B. Taghezouit and Y. Sun, “Robust and flexible strategy for fault detection in grid-connected photovoltaic systems,” *Energy Conversion and Management*, vol. 180, pp. 1153–1166, 2019.

- [14] A. Mellit, "Development of an expert configuration of standalone power PV system based on adaptive neuro-fuzzy inference system (ANFIS)," in *MELECON 2006–2006 IEEE Mediterranean Electrotechnical Conf.*, Malaga, Spain, pp. 893–896, 2006.
- [15] B. Taghezouit, F. Harrou, Y. Sun, A. H. Arab and C. Larbes, "A simple and effective detection strategy using double exponential scheme for photovoltaic systems monitoring," *Solar Energy*, vol. 214, pp. 337–354, 2021.
- [16] A. Bakdi, W. Bounoua, A. Guichi and S. Mekhilef, "Real-time fault detection in PV systems under MPPT using PMU and high-frequency multi-sensor data through online PCA-KDE-based multivariate KL divergence," *International Journal of Electrical Power & Energy Systems*, vol. 125, pp. 106457, 2021.
- [17] J. A. Reyes-Malanche, F. J. Villalobos-Pina, E. Cabal-Yepez, R. Alvarez-Salas and C. Rodriguez-Donate, "Open-circuit fault diagnosis in power inverters through currents analysis in time domain," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–12, 2021.
- [18] S. Xu, S. Tao, W. Zheng, Y. Chai, M. Ma and L. Ding, "Multiple open-circuit fault diagnosis for back-to-back converter of pmsg wind generation system based on instantaneous amplitude estimation," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–13, 2021.
- [19] T. Kamel, Y. Biletskiy and L. Chang, "Real-time diagnosis for open-circuited and unbalance faults in electronic converters connected to residential wind systems," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 3, pp. 1781–1792, 2016.
- [20] Z. Huang, Z. Wang and H. Zhang, "Multiple open-circuit fault diagnosis based on multistate data processing and subsection fluctuation analysis for photovoltaic inverter," *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 3, pp. 516–526, 2018.
- [21] G. Yating, W. Wu, L. Qiongbina, C. Fenghuang and C. Qinqin, "Fault diagnosis for power converters based on optimized temporal convolutional network," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–10, 2021.
- [22] M. Halwachs, L. Neumaier, N. Vollert, L. Maul, S. Dimitriadis *et al.*, "Statistical evaluation of PV system performance and failure data among different climate zones," *Renewable Energy*, vol. 139, pp. 1040–1060, 2019.
- [23] L. B. Bosman, W. D. Leon-Salas, W. Hutzler and E. A. Soto, "PV system predictive maintenance: Challenges, current approaches, and opportunities," *Energies*, vol. 13, no. 6, pp. 1398, 2020.
- [24] Z. Usman, J. Tah, H. Abanda and C. Nche, "A critical appraisal of pv-systems' performance," *Buildings*, vol. 10, no. 11, pp. 192, 2020.
- [25] K. V. S. Bharath, A. Haque and M. A. Khan, "Condition monitoring of photovoltaic systems using machine learning techniques," in *2018 2nd IEEE Int. Conf. on Power Electronics, Intelligent Control and Energy Systems (ICPEICES)*, Delhi, India, pp. 870–875, 2018.
- [26] K. Dadhich, V. S. B. Kurukuru, M. A. Khan and A. Haque, "Fault identification algorithm for grid connected photovoltaic systems using machine learning techniques," in *2019 Int. Conf. on Power Electronics, Control and Automation (ICPECA)*, New Delhi, India, pp. 1–6, 2019.
- [27] V. S. B. Kurukuru, F. Blaabjerg, M. A. Khan and A. Haque, "A novel fault classification approach for photovoltaic systems," *Energies*, vol. 13, no. 2, pp. 308, 2020.
- [28] S. Silvestre, A. Chouder and E. Karatepe, "Automatic fault detection in grid connected PV systems," *Solar Energy*, vol. 94, pp. 119–127, 2013.
- [29] S. Ali, M. S. Khan, V. S. B. K., M. A. Khan and A. Haque, "Failure rate basics for a case study on grid connected photovoltaic plant," in *2019 Int. Conf. on Power Electronics, Control and Automation (ICPECA)*, New Delhi, India, pp. 1–6, 2019.
- [30] L. M. Moore and H. N. Post, "Five years of operating experience at a large, utility-scale photovoltaic generating plant," *Progress in Photovoltaics: Research and Applications*, vol. 16, no. 3, pp. 249–259, 2008.
- [31] W. Chine, A. Mellit, A. M. Pavan and S. A. Kalogirou, "Fault detection method for grid-connected photovoltaic plants," *Renewable Energy*, vol. 66, pp. 99–110, 2014.

- [32] X. Zhang, W. Zhang, W. Sun, X. Sun and S. Kumar Jha, "A robust 3-D medical watermarking based on wavelet transform for data protection," *Computer Systems Science and Engineering*, vol. 41, no. 3, pp. 1043–1056, 2022.
- [33] X. Zhang, X. Sun, X. Sun, W. Sun and S. Kumar Jha, "Robust reversible audio watermarking scheme for telemedicine and privacy protection," *Computers, Materials and Continua*, vol. 71, no. 2, pp. 3035–3050, 2022.
- [34] A. C. Rencher and W. F. Christensen, in *Methods of Multivariate Analysis*, 3rd ed., vol. 33, no. 3, Hoboken, New Jersey, United States, Wiley, 2012.
- [35] P. de and B. Harrington, "Sigmoid transfer functions in backpropagation neural networks," *Analytical Chemistry*, vol. 65, no. 15, pp. 2167–2168, 1993.
- [36] R. J. de Andrade Vieira and M. A. Sanz-Bobi, "Failure risk indicators for a maintenance model based on observable life of industrial components with an application to wind turbines," *IEEE Transactions on Reliability*, vol. 62, no. 3, pp. 569–582, 2013.
- [37] H. M. Fekry, A. A. Eldesouky, A. M. Kassem and A. Y. Abdelaziz, "Power management strategy based on adaptive neuro fuzzy inference system for AC microgrid," *IEEE Access*, vol. 8, pp. 192087–192100, 2020.
- [38] M. J. Usmani and A. Haque, "Power management of solar pv systems for peer load," in *2020 IEEE Int. Conf. on Power Electronics, Smart Grid and Renewable Energy (PESGRE2020)*, Cochin, India, pp. 1–6, 2020.
- [39] M. A. Khan, A. Haque and V. S. Bharath, "Dynamic voltage support for low voltage ride through operation in single-phase grid-connected photovoltaic systems," *IEEE Transactions Power Electron.*, vol. 36, no. 10, pp. 12102–12111, 2021.
- [40] M. A. Khan, A. Haque, V. S. B. Kurukuru and S. Mekhilef, "Advanced control strategy with voltage sag classification for single-phase grid-connected photovoltaic system," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 3, no. 2, pp. 258–269, 2022.