The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH PAPER

Deep-learning–based method for faults classification of PV system

Sayed A. Zaki[1,2] | **Honglu Zhu**[1,3] | **Mohammed Al Fakih**[1] | **Ahmed Rabee Sayed**[1,2] | **Jianxi Yao**[1]

[1] School of New Energy, North China Electric Power University, Beijing, China

[2] Faculty of Engineering, Cairo University, Giza, Egypt

[3] The State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources, North China Electric Power University, Beijing, China

Correspondence
Sayed A. Zaki, School of Renewable and Clean Energy, North China Electric Power University, 102206, Beijing, China.
Email: eng_sayed_002010@yahoo.com

Abstract

The installation of photovoltaic (PV) system, as a renewable energy source, has significantly increased. Therefore, fast and efficient fault detection and diagnosis technique is highly needed to prevent unpredicted power interruptions. This is obtained in this study in the following steps. First, an efficient meta-heuristic algorithm is proposed for extracting the optimal five parameters of the PV model in order to assist the MATLAB simulation model. It is used due to its simplicity and high efficiency in building the PV array simulation. Second, a new PV system deep-learning convolutional neural network (CNN) fault classification method is presented for the advantage of automatic feature extraction, which reduces the computational burden and increases the high classification capability. Finally, for the practical and theoretical validation of the employed CNN model, normal and six fault cases are selected based on different atmospheric conditions. At same time, three electrical indicators are analysed and accordingly chosen as inputs to the proposed classification model. Moreover, the proposed model is compared with other machine-learning models.

1 | INTRODUCTION

Renewable sources of energy recently have noticeable donation to sustain sources of power generation. This is due to the yearly rise in energy demand, the bad impact of non-sustainable energy sources as well as global degradation of fossil fuel such as coal and oil. Among these renewable sources, solar energy is one of the most common widely utilised power sources due to its friendly environmental impact, zero-running cost, and simple technical issues needed for efficient operation. Currently, photovoltaic (PV) modules, as the main part of solar energy, is growing rapidly all over the world due to the significant progress of the developed technology as it helps to decrease the solar cells material costs [1]. However, the need for efficient fault diagnosis techniques has increased also for better monitoring the outdoor-installed PV systems considering the variation of environmental conditions.

For effective fault detection methods, modelling the PV system mathematically plays an important key on the accuracy of the classification technique. This is because it has a remarkable role in obtaining the optimal parameters, design, and assessment of the PV solar system fault diagnosis methods [2, 3]. Although the manufacturers of solar modules provide the electrical characteristics of PV modules at standard test condition (STC) to users to help in building the simulation model and other calculations, these characteristics can change from its nominal values under actual long-term working condition. Moreover, the PV system operation depends on the environmental conditions such as temperature and solar irradiance [4]. Therefore, an efficient PV model using measured current-voltage (I-V) data is needed to accurately and efficiently optimise the PV parameters [5]. There are many models that are formulated to characterise the I-V curve. The most commonly used for PV modelling is the single diode, double diode, and the PV module models [6]. On the other hand, an efficient PV optimisation algorithm is needed to extract the electrical parameters from PV models. Recently, several algorithms have been dedicated that can be classified into three sets: Analytical algorithms [7], deterministic algorithms [8], and meta-heuristic algorithms [9]. Based on the latter set, the improved teaching-learning-based optimisation algorithm (ITLBO) [10] is used in this study for extracting the electrical parameters from PV module to be employed in the

simulation model due to its simplicity in execution as well as no special requirements.

In general, studying the different faults, in case of low-solar irradiance level, is computationally challengeable for the PV fault diagnosis because disturbances introduced on the different indicators might be unnoticeable. Valuable studies [11–13] offer several algorithms to detect PV faults such as low-mismatch faults, and line-to-line fault (LL) under low-solar irradiance levels. However, such studies provide low accuracy in fault identification and need high-cost measurement sensors. In this work, the low irradiance is taken into consideration when studying the different faults for achieving better performance of the proposed model.

Nowadays, the artificial intelligence (AI) techniques are commonly used for the diagnosis of PV fault types due to high classification accuracy. Artificial neural network (ANN), probabilistic neural network (PNN), and fuzzy c-means (FCM) are lately used in a wide range in [14–16]. However, these AI techniques have some demerits such as huge data amount and computational time for training stage, sensitive for small environmental condition variations, and updating the collected data periodically are needed due to degradation ageing problems in solar PV modules. Furthermore, the advantage of monitoring the operation of PV systems gives the user a huge data amount for detecting different known and unknown faults. Consequently, using this kind of data-based models is relevant, and the process of features extraction is effective for PV fault diagnosis issues. For the purpose of extracting the desired features, convolutional neural network (CNN) as a multivariate model is recently used for data image representation and also for faults classification task.

In the case of CNN-based classification model, CNN utilises mainly in image recognition for its high accuracy in classification. Moreover, using CNN in monitoring numerous engineering schemes are increased for the advantage of automatic features extraction process. In the field of detecting the motor faults, Lee et al. [17] use signal databases containing a univariate with bivariate data time series, then, they implement CNN as a feature extraction and classification technique to raise the accuracy of bearing fault classification rate. Chen et al. [18] suggest CNN for the classification and extract-features technique. This is obtained by dividing the collected multivariate time-series data into univariate data for each parameter, then, the feature extraction process was performed for each parameter individually. In case of PV solar cells, Li et al. [19] conduct one dimensional CNN to classify the different kinds of PV module defects such as yellowing, dust-shading, and corrosion of gridline using aerial images in large-scale PV plants. However, the equipment used in the work is expensive, and the CNN implemented only on the offline operating condition.

In this study, the main innovations are trifold:

1. Based on meta-heuristic techniques, the ITLBO is advised to extract the electrical parameters of PV modules for the simulation model.
2. The CNN fault classification technique is proposed to achieve high performance of the fault diagnosis tasks, considering the advantage of automatic features extraction from
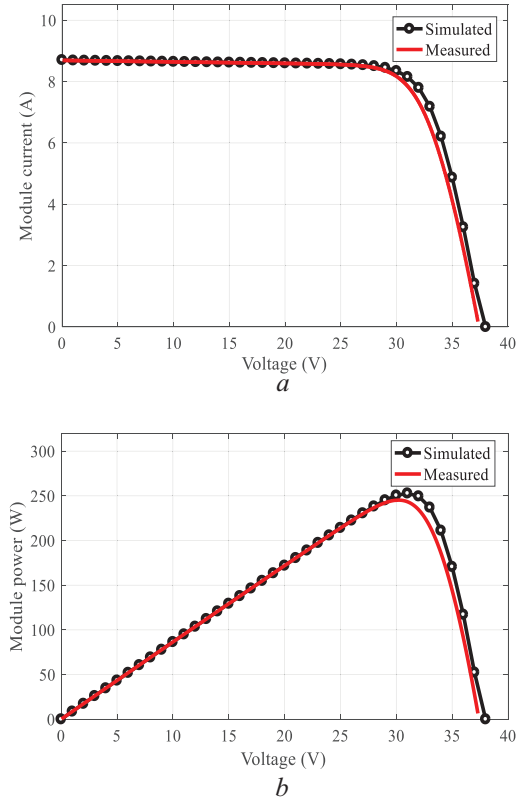

a

b

FIGURE 1 The measured curves and simulated ones of the photovoltaic (PV) model obtained by improved teaching-learning–based optimisation (ITLBO) algorithm at 1000 W/m² irradiance and 25°C temperature: **(a)** The current-voltage (*I–V*) curves, **(b)** The *P–V* curves

input datasets, as softmax layer, to obtain the classification output result.
3. The effectiveness of the proposed technique is validated by simulated and experimental case studies. In addition, it is also compared with three benchmark AI models, namely, two-stage support vector machine (SVM) [13], ANN [14], and PNN [15].

2 | PV SYSTEM MODEL AND DESCRIPTION OF TYPICAL FAULT INDICATORS

2.1 | PV modelling

The equivalent circuit of the PV module is shown in Figure 1. The PV module can be modelled by connecting some diodes in series representing the series connected solar cells, and some diodes in parallel representing the parallel branches of solar cells (in this study, each PV module has 60 cells connected in series and no parallel branches of solar cells), then the generated current can be written in Equation (1) as [6]

$$I = I_{ph} - I_O \left(e^{\frac{V + I N_s R_s}{n N_s V_{th}}} - 1 \right) - \left(\frac{V + I N_s R_s}{N_s R_{sh}} \right) \quad (1)$$
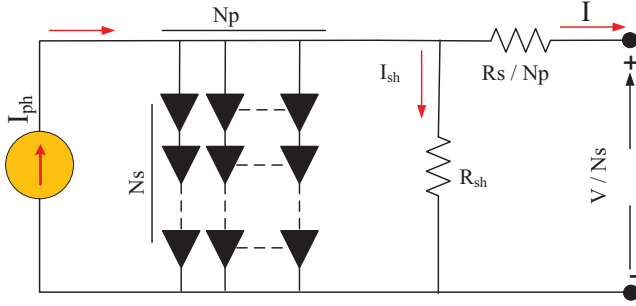
FIGURE 2 The model circuit of the PV module

TABLE 1 Electrical parameters of the PV module at STC

Parameter	Value
$I_{m,mod}$ (A)	8.12
$V_{m,mod}$ (V)	30.2
P_{max} (W)	245
$I_{sc,mod}$ (A)	8.69
$V_{oc,mod}$ (V)	37.4
α (A/K)	0.0006
B(V/K)	–0.0031

where the thermal voltage V_{th} which can be expressed as in Equation (2)

$$V_{th} = \frac{nKT}{q} \qquad (2)$$

To model an accurate PV module, five parameters are required to be mathematically calculated. These parameters are I_{ph}, I_O, R_s, R_{sh}, and n. The main reason for extracting these parameters is to get the minimum absolute error between the measured with the simulated data of the generated current.

The ITLBO algorithm presented in [10] is used in this study as an accurate optimisation technique for parameter extraction of the PV module. Compared with the existing teaching-learning–based algorithms, the proposed algorithm has two main improvements, namely, (i) in the teacher stage, a new teaching plan is proposed where the learners are divided into two different groups based on the learning level, then the best learners are guided from their corresponding teacher as well as themselves; and (ii) in the learner stage, instead of only one learner to exchange knowledge to another learner, the new plan of ITLBO is to make the well-learners learn from the other two learners for accurate search path, while the poorer learners learn from the knowledge of four different learners in order to increase the global search capability and provide the population variety. The simulated data obtained by the ITLBO algorithm at STC operating condition is compared with the measured values in Figure 2. It is observed that the simulated data is almost similar to the measured one which provides accurate modelling for the PV module to get the five parameters The available electrical parameters of the PV module provided by the manufacturer during the optimisation process are listed in Table 1, while the obtained five parameters can be shown in Table 2.

2.2 | Indicators selection

Selecting the indicators for detecting the faults in PV system must satisfy the following aspects: (i) Ability of identifying and discriminating the different kinds of fault under the variation of solar radiation and module temperature, (ii) suitable for dif-

TABLE 2 The calculated five parameters of PV module at STC

Optimal parameter	Value
I_{ph}(A)	9.33
I_o(A)	10^{-10}
R_s($'\Omega$)	0.24
R_{sh}($'\Omega$)(V)	593.24
n(p.u.)	1.102

ferent PV systems scale and configurations, (iii) ability of using optimum number of indicators to help accelerate the classification operation and minimise the used memory, (iv) measuring and building the selected indicators must be simple and easy for implementation. Therefore, in this study, three indicators are implemented from the whole PV array rather than calculating those for each string or module.

Voltage, current, and power, which are obtained from the I-V and P-V curves, are the three common indicators used in identifying different faults of the PV array, see examples in [14–16, 20, 21]. Although, these indicators could discriminate between the healthy the faulty conditions, implementing these indicators, in our study, has made an overlapping between the different selected faults for the irradiance variation. More specifically, every two data points have the same working situation and voltage levels attained at some fault cases. Therefore, for efficient fault diagnosis method that has a better sense to environmental condition variation and fault happening, three indicators are chosen in this study for analysing the characteristics of different faults. The fault diagnosis procedure is based on extracting these indicators under the variation of solar irradiance and module temperature. Furthermore, analysing the selected indicators in the PV system represents the key role to detect and accurately classify the different faults.

The selected indicators for the PV array are *Normalised* V_m, *Normalised* I_m, and *FF*, and their expressions can be obtained as in Equations (3) to (5), respectively [11, 22].

$$Normalised \ V_m = \frac{V_m}{V_{oc,ref}} \qquad (3)$$

$$Normalised\ I_m = \frac{I_m}{I_{sc,ref}} \quad (4)$$

$$FF = \frac{V_m * I_m}{I_{sc} * V_{oc}} \quad (5)$$

To calculate these indicators, the electrical parameters of the PV array, namely, V_m, I_m, I_{sc} and V_{oc}, can be determined analytically at any solar irradiance S and temperature T as in Equations (6) to (9), respectively [23].

$$V_m = N_{s,mod} * \left[V_{m,STC}\left[1 + \alpha\left(T - T_{STC}\right)\right] + V_{th}*ln\left(\frac{I_{sb}}{I_{sb,STC}}\right)\right] \quad (6)$$

$$I_m = N_{P,mod} * \left[I_{m,STC}\left(\frac{S}{S_{STC}}\right)\left[1 + \alpha\left(T - T_{STC}\right)\right]\right] \quad (7)$$

$$I_{sb} = N_{P,mod} * \left[I_{sb,STC}\left(\frac{S}{S_{STC}}\right)\left[1 + \alpha\left(T - T_{STC}\right)\right]\right] \quad (8)$$

$$V_{oc} = N_{s,mod} * \left[V_{oc,STC}\left[1 + \beta\left(T - T_{STC}\right)\right] + V_{th}*ln\left(\frac{I_{sb}}{I_{sb,STC}}\right)\right] \quad (9)$$

2.3 | Typical faults occurring in PV arrays

There are various types of faults that may occur in the PV system. Some of faults on the DC side that frequently occur have excessive power losses and reduction in the efficiency of the PV system, including short-circuit (SC) faults, LL faults, open circuit (OPEN), partial shading (Shad), and degradation. It is noted that the most commonly occurring faults in the existing studies are LL, OPEN and Shad faults [13–16].

The LL faults (or bridging faults) are conducted here by generating an SC connection between two points in a string or among different PV strings. Practically, this fault type may have occurred due to either water entering to the conductors, chewing by animals, mechanical damage of insulation, or junction box damage by DC arcs. LL faults can cause fire damage in large-scale PV systems. Moreover, they may generate reverse fault current, with its value depending on the voltage difference among the two faulty points. Low-voltage difference would provide a low fault currents, and this becomes a challenge to detect this LL fault cases using traditional fault detection methods. In addition, in such LL fault cases, the protection devices may fail to sense such currents (see [13, 24] for more details). Therefore, to address these concerns, this study focuses on diagnosing LL1 and LL2 fault cases under the variation of atmospherical conditions, particularly in low-irradiance levels. For implementing the LL fault in this study, two fault scenarios, denoted as LL1 and LL2, represent four and six short-circuited modules mismatch between two strings, respectively. However, such fault has not commonly occurred in the PV arrays, and

their effect should be discussed and detection techniques should be utilised to prevent probable power losses or fire damage. As indicated by [11], the resistance of SC can be assumed to equal zero, which is consistent with our setting for faults LL1 and LL2.

Shading faults are recently modelled and studied due to its distinct impact on the power output from the PV systems during the day operation. Although, their impacts are not permanent, they should be distinguished from the other fault types for the exact classification in order to serve the protection devices. In this study, two fault scenarios of shading faults, denoted as shad1 and shad2, are selected to be examined. These faults represent four- and six-shaded modules in same string, respectively. Note that the shunt bypass diode is assumed to normally operate for the shaded modules. The shading factor is set at 50%, which means only half of the full solar irradiance is received by the module surface, while the temperature is same as STC. However, the shading fault can be performed at different patterns and shading factors in the PV array. This study considers that the shading is almost fixed over the day, that is, the constant shading factor is applied. More specifically, in small residential PV arrays, the physical area is relatively small and the solar irradiance can be assumed constant on all PV modules over the day [25]. Moreover, in such arrays, the shading faults are assumed to occur due to the nearby buildings or the associated equipment. Therefore, the same numbers of PV-shaded modules may be affected by this small moving shadow along the day, wherever the module location is, inside the examined string. So, the number of shaded modules are considered fixed along the daytime, as similar treatment is found in [15, 16].

Finally, OPEN circuit faults are modelled by disconnecting electric wiring between the PV modules. Two faults are considered, namely, OPEN1 and OPEN2, to denote the OPEN circuit faults in one and two PV strings, respectively. It should be noted that every fault class is simulated separately, that is, the study of multiple faults occurring at the same time is not included in this study.

The PV array under study contains three strings, where each string consists of 13 modules connected in series as shown in Figure 3(a). Each module is connected in parallel with shunt bypass diode to avoid hotspots formation and power loss in shading condition as shown in Figure 3(b). Moreover, seven experimental operating tests on the DC side have been simulated to evaluate the fault diagnosis method; one normal condition, and six fault cases can be seen in Figure 4 (LL1, LL2, OPEN1, OPEN2, Shad1, and Shad2). The different typical faults are simulated using MATLAB Simulink in case of STC as shown in Figure 5. According to Figure 5, in the LL1 and LL2 fault cases, V_{oc} and V_m are declined, while the I_{sc} and I_m are unchanged compared with the normal state. On the other hand, in the OPEN1 and OPEN2 fault cases, the value of I_{sc} and I_m are significantly decreased, where the V_{oc} and V_m are almost the same. Additionally, in Shad1 and Shad2 fault cases, the I_m and V_m, are significantly affected, while I_{sc} and V_{oc} are similar with those of the normal condition.
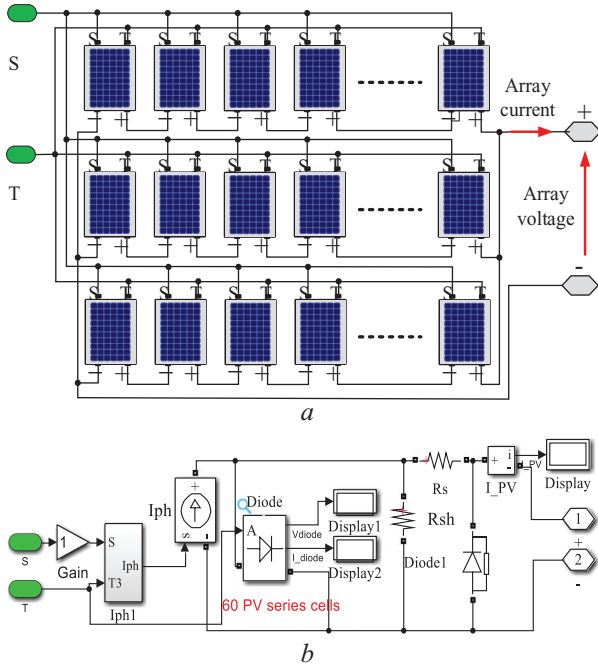
FIGURE 3 (a) Schematic diagram of the PV array with 3×13, and (b) the equivalent circuit model of the PV module
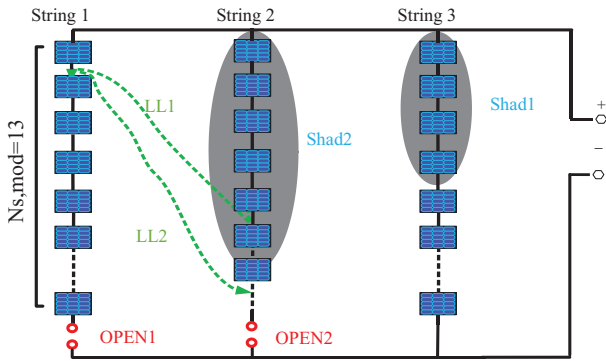

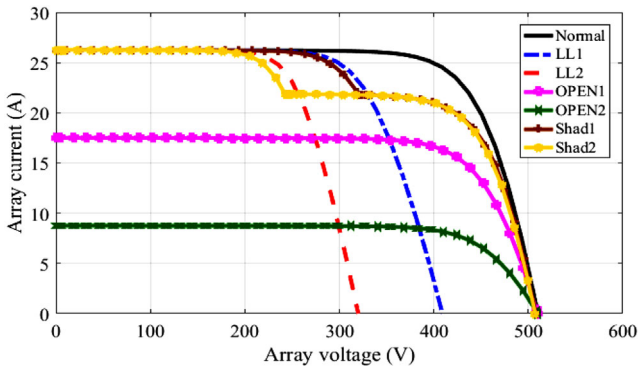**FIGURE 4** The DC side of the PV array with the examined six fault cases

FIGURE 5 The I–V curves of the PV array at one normal and six faults cases at standard test condition (STC)

3 | METHODOLOGY

First, the analysis of the selected three indicators is performed. Then, a fault detection and diagnosis method is presented based on the deep-machine model of the standard CNN.

3.1 | The analyses of selected indicators for typical faults

To perform indicators analysis, the typical faults are conducted under different environmental conditions, that is, variable temperature and solar irradiance in three consecutive days in June. A wide range of irradiance from 5 to 1000 W/m^2 and temperature data samples from 5 to 55°C are chosen with a time step of 15 min between every two samples with solar irradiance values as shown in Figure 6(a), where all samples with zero irradiance, that is, during the night, are eliminated.

To simulate the shading faults, the solar irradiance is multiplied by gain equals 0.5 to be feed to the shading modules, while the temperature is unchanged as in normal condition. Then, the selected indicators, *Normalised* V_m, *Normalised* I_m, and *FF*, are calculated using Equations (3) to (5) as shown in Figures 6(b) to (d), respectively.

From the presented Figure 6(b), it is observed that the *Normalized* V_m significantly declined in cases LL1 and LL2. For *Normalised* I_m, it decreased in OPEN1 and OPEN2 cases as shown in Figure 6(c). Moreover, the third indicator *FF* is influenced and declined in cases Shad1 and Shad2 as shown in Figure 6(d); the reason is that these fault types have a direct impact on V_m and I_m (refer Figure 4). Based on the aforementioned discussions, the selected three indicators are mainly and differently influenced by the fault type, which, in result, indicates the ability of the proposed methodology to detect and classify different kind of faults. Additionally, this would help to decrease overlapping between the collected data samples, and hence, raise the proposed method efficiency. Finally, the conclusion of the three indicators distribution is summarised in Table 3.

3.2 | CNN structure

CNN is commonly utilised in image recognition and classification. It has been successfully used to identify faces, different objects and so forth. Therefore, CNN is an important machine-learning tool for most practitioners today.

The one-dimensional CNN, which is used in this study, consists of two consecutive steps [26] as illustrated in Figure 7:

1. The first step is the feature extraction which contains three stages: The input layer, convolutional, and pooling stages. The convolutional layer is to extract the features across the raw data of the input layer by a filtre and stores them as a con-
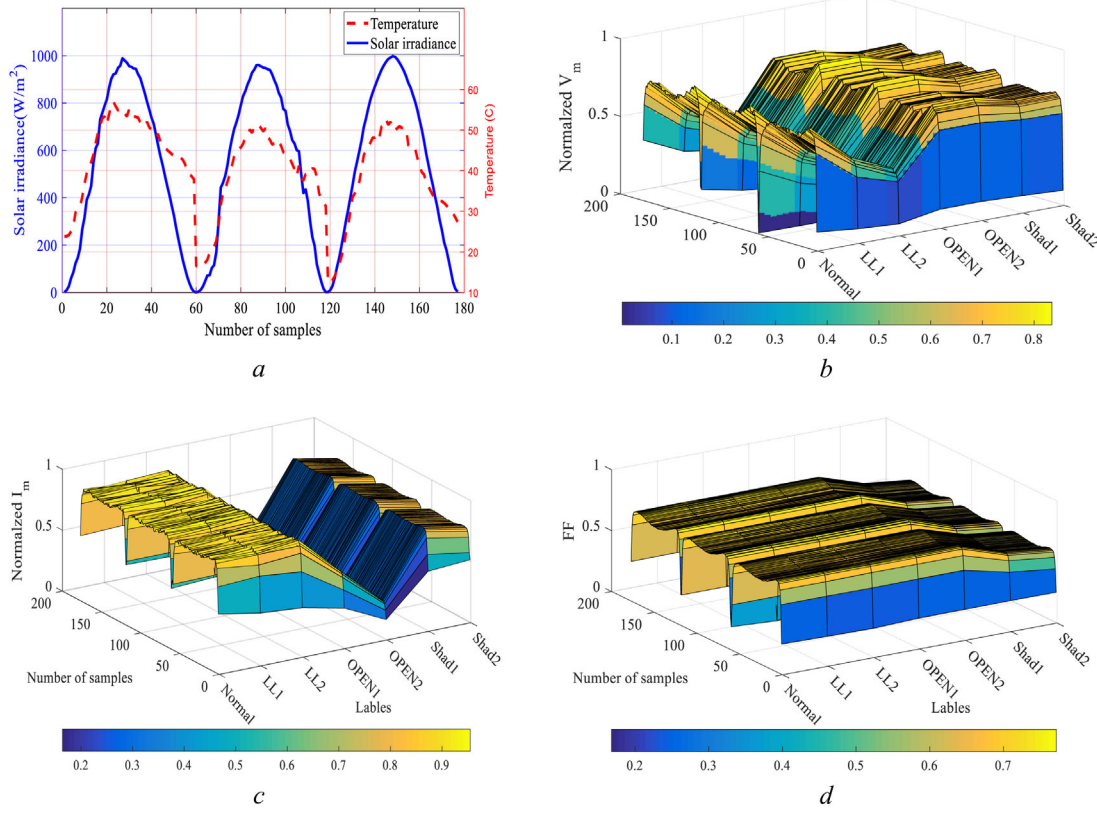
FIGURE 6 The distribution of the selected indicators in normal and different fault conditions in three days. (a) The environmental conditions of module irradiance and temperature, (b) the *Normalised V_m*, (c) the *Normalised I_m*, and (d) the *FF*

volved feature, which can be conducted using sigmoid activation function. The pooling layer function is to reduce the size of the convolved feature, which progressively reduces the computational time in the network. Max-pooling is the method that usually utilised to complete the process of the pooling layer.

2. The second step is classification, which contains the fully connected and the output stages, where the inputs of fully connected stage are the features attained from the pooling stage, and then classification task using softmax in the output stage is performed.

The samples dataset is collected from the PV simulation model under normal and abnormal conditions. This dataset represents a three-dimensional matrix for data of the three indicator samples calculated in Section 3.1.

3.3 | The proposed fault diagnosis method

A feature map is conducted based on a one-dimensional structure, where the pooling layer is used to reduce the dimension of the conventional layer at every individual feature map. During the training process of the CNN, weights are updated using the gradient descent method to minimise the loss (classification error) function value. This implies that the larger magnitude of a

TABLE 3 The indicators variation of PV array in case of normal and different fault cases

Faults	Normalised V_m	Normalised I_m	Fill factor (*FF*)
Normal	0.68–0.8	0.9–0.95	0.7–0.75
LL1	**0.5–0.54**	–	–
LL2	**0.38–0.44**	–	–
OPEN1	–	**0.6–0.62**	–
OPEN2	–	**0.3–0.31**	–
Shad1	0.68–0.73	–	**0.6–0.63**
Shad2	0.69–0.74	–	**0.56–0.58**

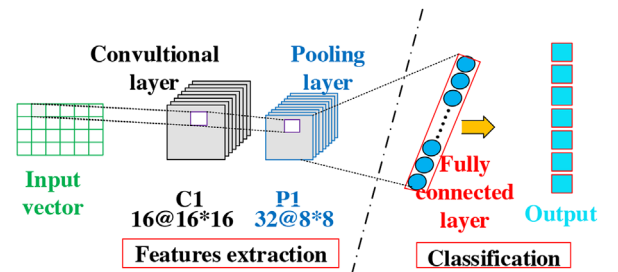Note: (-) means no change compared with normal case.

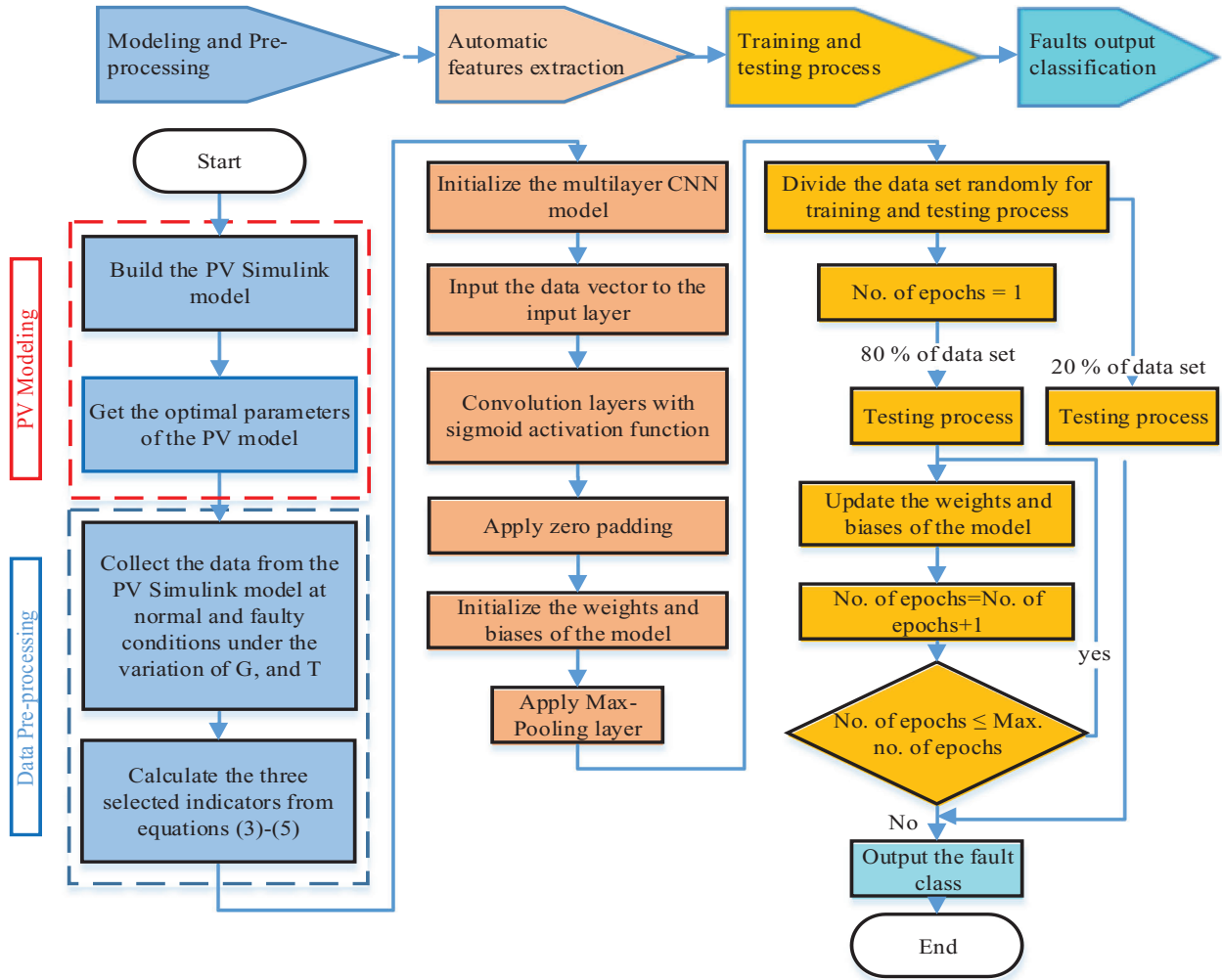**FIGURE 7** The proposed convolutional neural network (CNN) architecture

FIGURE 8 The proposed CNN-based model for fault classification

given weight, to the extracting features significant for classification. Also, zero-padding is applied to prevent shrinking as well as information loss happening at the borders. The flowchart of applying the proposed CNN model can be shown in Figure 8.

4 | RESULTS ANALYSES

The proposed model is validated in this section theoretically and physically for assessing the classification accuracy.

4.1 | Simulation tests

In order to validate the proposed method, a case study is performed in this part using MATLAB Simulink software. The data samples distribution between the three indicators are collected in three-dimensional matrix for generating the input of the proposed CNN model. The data is gathered as shown in Figure 9 as described previously in Section 3.1. This data will

follow, first, the pre-processed operation for data filtering, elimination of noise and so forth. Then, the dataset are separated for two stages: The training and testing stages for diagnosing the CNN model. The total dataset consists of $1239 = (177*7)$ including normal and six different fault cases, each sample has the three selected input indicators.

4.1.1 | Training results

The dataset at the training stage is assessed by using 80% of the total dataset (118 dataset samples for each normal and fault cases). To perform the training process, the CNN technique was trained using the selected parameters shown in Table 4. Grid search technique is conducted to get these parameters by trying some combinations to the best one suitable for the training dataset.

One of the key points for preventing the early convergence process for better model performance is to choose the suitable optimisation algorithm, which helps to increase the training computational speed and the accuracy of fault
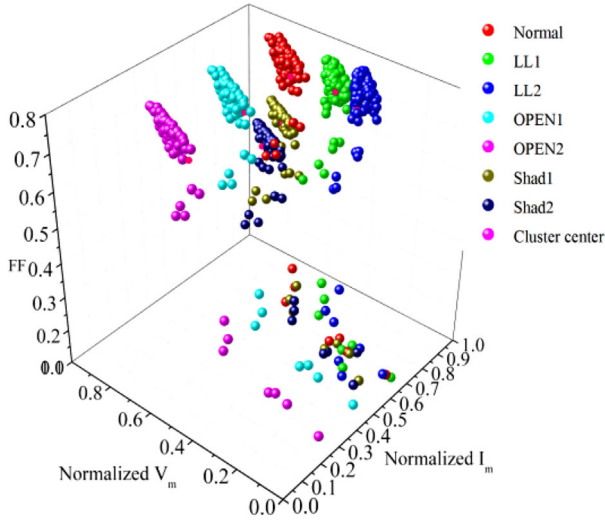
FIGURE 9 The data samples distribution of the three indicators

TABLE 4 The parameters of the proposed CNN

Parameters	Value
Input	25
Feature input dimension	16
Batch dimension	128
Learning rate	0.01
RMSprop optimiser variables	
No. of epochs	50
Input layer nodes	32
Fully connected layer nodes	40
Softmax layer nodes	6

TABLE 5 The confusion matrix of CNN classification model for the training samples

Case	Normal	LL1	LL2	OPEN1	OPEN2	Shad1	Shad2
Normal	98.23	0	0	0	0	1.77	0
LL1	0	98.6	1.4	0	0	0	0
LL2	0	1.66	98.34	0	0	0	0
OPEN1	0	0	0	100	0	0	0
OPEN2	0	0	0	0	100	0	0
Shad1	1.04	0	0	0	0	96.91	2.05
Shad2	1.45	0	0	0	0	1.5	97.05

TABLE 6 The confusion matrix of CNN classification model for the testing samples

Case	Normal	LL1	LL2	OPEN1	OPEN2	Shad1	Shad2
Normal	99.1	0	0	0	0	0.9	0
LL1	0	98.69	1.31	0	0	0	0
LL2	0	1.24	98.74	0	0	0	0
OPEN1	0	0	0	100	0	0	0
OPEN2	0	0	0	0	100	0	0
Shad1	1.04	0	0	0	0	97.41	2.55
Shad2	1.45	0	0	0	0	1.5	97.05

4.1.2 | Testing results

The testing process is developed to measure the performance and accuracy of the proposed technique. The remaining total dataset of 20% is used to test the proposed method. These data samples are gathered in the test dataset including the three typical indicators. All the test cases are accurately detected and categorised in the relevant case. Also, the proposed CNN method could distinguish successfully between the normal case and different faulty ones. These results are obtained in Table 6, which shows the confusion matrix of the test dataset with the corresponding cases. The observations show that the average classification accuracy is 98.6% and with few data samples are wrongly classified into different cases: Around 0.9 % for normal condition case classified inaccurately as Shad1 fault case, 1.31 % for fault case LL1, and 1.24% for fault case LL2. Furthermore, in case of Shad1 and Shad2, 2.55% and 1.5% are misclassified to wrong fault case as shown, with 100% classification accuracy in case of OPEN1 and OPEN2 because of large Euclidean distance between these cases and the others. These observations approve that the CNN method can accurately detect and classify the typical fault types. Moreover, the proposed method is efficiently robust under noise condition, however, using high precision measurement sensors in real-time monitoring can raise the classification accuracy and improve the proposed method performance.

classification in the CNN model. As stated in [27], there are three types of the commonly used algorithms, namely, RMSprop, Adam and Adadelta. In this study, the RMSprop algorithm is implemented to obtain the learning rate, batch size, and number of epochs related to the technical implementation. The optimisation implementation is obtained based on calculating the mean data of the nearest amount of the grade weight. It is useful to process the dynamic data variables such as the current, and voltage data at normal and faulty conditions. For better monitoring and analysis of the training process, the training accuracy and loss values between CNN model output and labelled ground truth are measured during the number of batches.

Furthermore, the results in Table 5 which show the confusion matrix of the training samples evaluate that a few data samples are misclassified in the training stage. Based on these efficient classification outcomes, the proposed method with the associated model gained from the training data samples can be used for predicting new fault types with average accuracy of 98.3%.
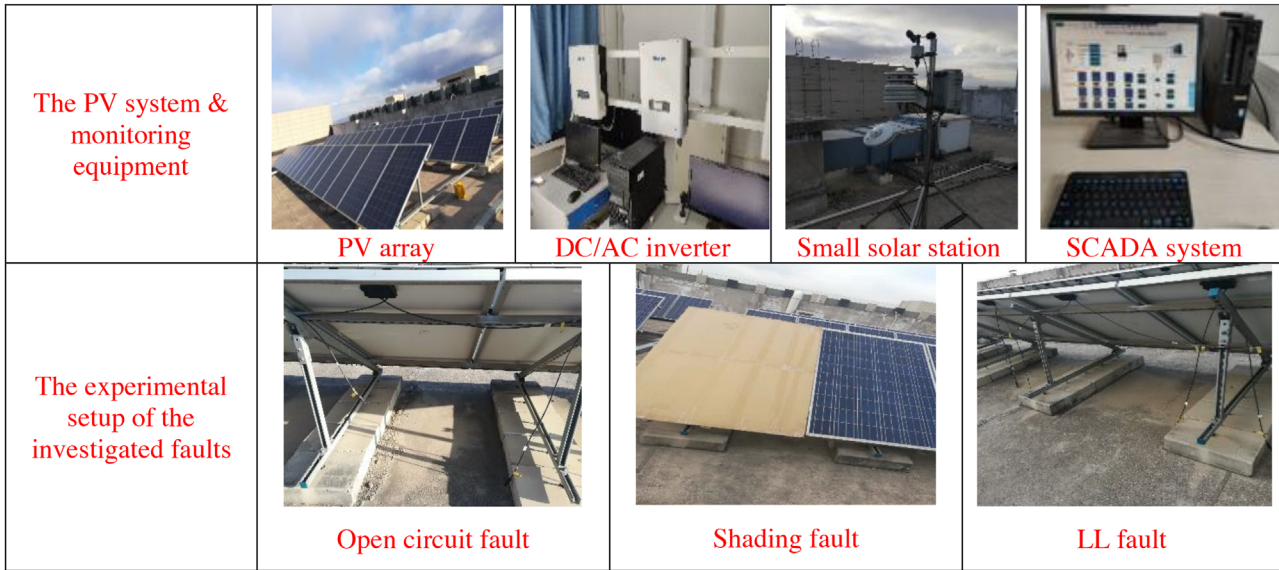
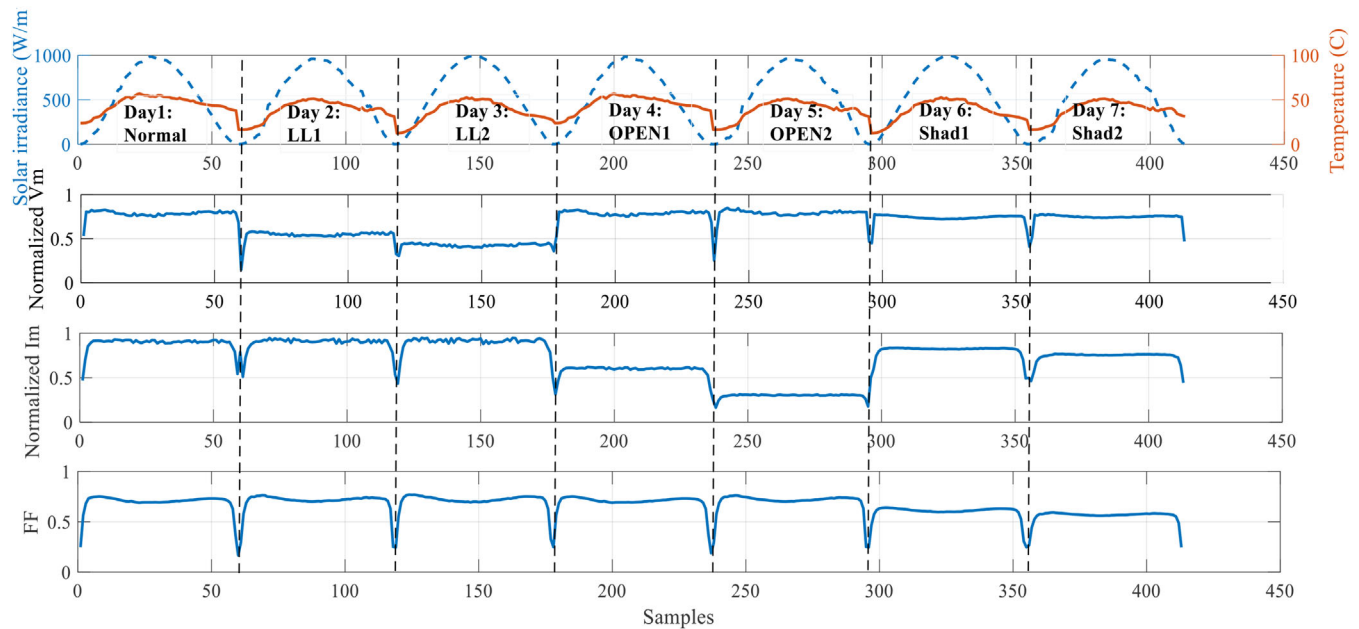**FIGURE 10** The overall PV system with the investigated faults

FIGURE 11 The data samples of the three indicators during the experimental tests

4.2 | Experimental tests

The proposed CNN model is implemented experimentally using the 9.56 kW PV array installed at the rooftop of North China Electric Power University. The overall PV system consists of three strings containing 13 JKM245p modules connected in series, where the electrical parameters of the module is previously presented in Table 1. In addition, the PV system is connected to AC inverter including AC and DC ammeter and voltammeter to the current, voltage, and power data. Also, the inclined surface irradiance and surface tempera-

ture are measured by the small solar station. For full supervision along the day, all the experiments are monitored by Supervisory Control and Data Acquisition (SCADA) system and high specification PC computer. The description of the system and examined faults setting can be shown in Figure 10. The different fault types are performed experimentally during seven consecutive days to examine one normal condition, and six fault types, where each type is performed along a day from 6:00 AM to 6:00 PM. The data samples of the three indicators are collected every 15 min step of full solar irradiance and surface temperature in June, as shown in Figure 11. It is noticed that, compared with

TABLE 7 The classification accuracy of the proposed model for the experimental tests

Stage	Training stage	Testing stage
Max.	98.11	98.65
Min.	95.41	96.17
Mean	96.76	97.41
St. dev.	0.781	0.882

summary results of the indicators variation in Table 3 during the simulated tests, the same impact is detected in case of the experimental tests. For implementing the shading fault, the shaded module is conducted by covering half of it by a physical solid opaque.

For checking the classification accuracy, similarly to simulation tests, the collected dataset is divided for performing the training and testing processes. Then, the classification accuracies for both processes are recorded in Table 7, including max, min, mean, and standard deviation (St. dev.) for the examined cases. From the results, the mean accuracies for the training and testing processes are decreased to be 96.76% and 97.41% compared with 98.3% and 98.6% in the simulation tests, respectively. The reason is that a misclassification problem might have occurred for unknown faults. Moreover, the measurement errors of ammeter and voltammeters may lead to reduce the proposed model accuracy.

4.3 | Comparison with traditional methods

In this section, we present a comparison between the proposed method and other three traditional methods. First, a fault classification method is built in [11] using normalised voltage and current of the I-V curve and graph-based semi-supervised learning (GBSSL) method. Although, it has high-classification accuracy, this method needs to measure the voltage and current from the practical PV array for implementation. In addition, only permanent faults are examined, that is, shading faults are not included. Second, a fault detection method based on pre-calculated thresholds to detect various PV fault case studies is presented in [28]. In this study, only partial-shading fault is detected based on determining the indicators of power and voltage ratios. However, the exact faulted module can be determined with help of the statistical t-test, the method accuracy depends on the sensors efficiency for measuring power and voltage values. Third, fault detection method is built based on the comparison between normal and faulty thresholds. These thresholds are calculated in ranges for different seven features, which are determined from the I-V curves measurement in different fault cases. Although, the measured signals noise may lead to wrong discrimination between different features, a prior threshold ranges for each fault are needed. The comparison is presented in Table 8.

From the results in Table 8, the advantage of our proposed model is that it is capable, theoretically and experimentally, to classify temporary faults (shading) as well as permanent faults (OPEN and LL) with average high accuracy (98.5% during the simulated tests). However, it is relatively complex in implementation and needs additional computational burden for the training process.

4.4 | Comparison with AI models

In order to effectively compare between the other AI models, we selected four common benchmark factors to obtain this assessment using the confusion matrix. For evaluating the four models, these benchmarks are calculated for one normal and six fault cases examined in this study. These benchmark factors are as follows: precision (P), recall (R), and F-measure (F_m). The values of P and R can be calculated as in Equation (10) [30].

$$P = \frac{TC}{TC + FC}, \; R = \frac{TC}{TC + FN} \qquad (10)$$

For calculating P and R in different models, true positive (TC) indicates the samples, which are correctly categorised as the actual; false correct (FC) indicates that the samples are incorrectly categorised as the actual; true negative (TN) means the samples do not belong to the class in prediction or actual; and false negative (FN) means the samples are incorrectly predicted but actually belong to the same class. From Equation (10), the P definition is the percentage of relevant samples to the total predicted samples in the same class. While R is the percentage of relevant samples in individual class to the total actual samples in the same class.

Then, generally when assessing some models, the F-measure F_m is utilised to comprehensively evaluate the models due to which the values of P and R are evenly biased and do not reflect the exact performance of the model. Therefore, the general formula to obtain F_m for non-negative real m based on the P and R values is shown in Equation (11). In other words, F_m represents the harmonic understanding of P and R, where the best value of F_m is 1 at perfect values of them. According to experimental evidence, we select m to be 0.3 to calculate F_m in all models as in Equation (11).

$$F_m = \frac{\left(1 + m^2\right) * P * R}{m^2 * P + R} \qquad (11)$$

The proposed CNN model is compared with three machine-learning models: SVM, PNN, and ANN to measure the classification capability using the simulation dataset. This comparison is shown in Table 9. For more details, first, a two stage SVM is presented in [13], where various fault cases of LL faults were investigated. Multiresolution-signal-decomposition is utilised to extract the different features of LL faults manually from the input data samples of voltage and current. However, this manual features extraction may create errors during the fault classification. Moreover, according to Table 9, the average fault classification P, R and F_m are obtained for LL fault cases in the range of 90 % using SVM. These are due to the

TABLE 8 The comparison with traditional methods

Ref.	Year	Method	No. of indicators	Permanent/temporary faults	Simulated /experimental validation	Accuracy	Limitations
[11]	2015	Measuring voltage and current with graph-based semi-supervised learning (GBSSL)*	2	Permanent	Experimental	100	Need initial label samples
[28]	2016	Analysis of power loss	2	Temporary	Experimental	Not defined	1. Pre-calculated threshold is needed 2. Detection accuracy depends on measured sensors
[29]	2017	I-V curves	7	Both	Experimental	Not defined	Pre-calculated threshold is needed
Proposed	2020	I-V curves with CNN	3	Both	Both	98.5	(1) Time needed for training stage (2) Relatively complex

TABLE 9 The comparison of the proposed CNN model with other models

Fault Cases	SVM [13]			PNN [15]			ANN [14]			Proposed CNN- based		
	P (%)	R (%)	F_m (%)	P (%)	R (%)	F_m (%)	P (%)	R (%)	F_m (%)	P (%)	R (%)	F_m (%)
Normal	96.7	95.5	96.6	100.0	84.5	85.6	92.2	93.7	92.3	98.2	98.5	98.2
LL1	90.4	91.8	90.5	100.0	100.0	100.0	88.1	87.5	88.1	97.4	98.1	97.5
LL2	89.9	91.5	90.0	100.0	100.0	100.0	90.4	88.7	90.3	97.5	99.0	97.6
OPEN1	98.4	97.3	98.3	100.0	100.0	100.0	90.4	89.2	90.3	100.0	100.0	100.0
OPEN2	98.5	98.1	98.5	100.0	100.0	100.0	87.5	85.4	87.3	100.0	100.0	100.0
Shad1	93.4	95.8	93.6	100.0	100.0	100.0	86.2	84.7	86.1	96.2	97.4	96.3
Shad2	97.4	95.8	97.3	100.0	100.0	100.0	87.4	85.4	87.2	98.4	97.8	98.4
Overall	96.5	96.5	96.5	100.0	97.8	97.9	88.9	87.8	88.8	98.2	98.7	98.3

low accuracy of LL fault classification at low-solar irradiance level.

Second, in [15], a PNN-based model is described for the fault diagnosis in the PV system. The technique performance is analysed for comparison based on implementing the typical fault types as shown in Table 8. Then, the classification accuracies of P, R, and F_m are obtained from the confusion matrices stated in the study. However, the PNN model has a good overall accuracy for typical fault classification, the R value in case of normal is the lowest at 84.5%, which means a lot of samples that actually belong to normal condition are incorrectly predicted as other fault classes. Therefore, the fault diagnosis technique sends wrong signal alarms to the operators and may lead to several power interruptions. In addition, it uses additional Gaussian kernel FCM method with the PNN model to obtain the optimal clusters for better representation of the data samples. Therefore, the classification accuracy of this study will be influenced by the method's efficiency.

Lastly, in [14], the ANN-based fault detection method is investigated based on the thresholds of the electrical characteristics of I-V and P-V curves. The measurement-sensors accuracy used in collecting the data samples has been taken into account for effective detection of the different faults. However, manual pre-calculated thresholds are required to classify the typical faults. In addition, the P, R and F_m of the ANN model is the lowest for classifying the typical faults with values 88.9%, 87.8%, and 88.8%, respectively.

After this comparison, we found that the P, R and F_m of the CNN model is 98.2%, 98.7%, and 98.3%, respectively, which are nearly higher than the other AI models as shown in Table 9.

Finally, the results of simulated, experimental, and comparison with the other techniques summarise that the CNN-based model as a deep-learning method succeeded in competition with the other methods for detecting and diagnosing the typical faults in PV system.

5 | CONCLUSION

One of the deep-learning models is employed in this study to enhance the classification accuracy for detecting different faults in DC side of the PV array, and to eliminate the errors due to extracting the different features manually in other algorithms. First, ITLBO optimisation algorithm is applied to obtain the best values of the PV simulation model parameters to enhance the system capability and efficiency. Second, three selected indicators are calculated and analysed in case of normal and different fault cases representing LL, open circuit, and shading faults under atmospherical condition variations.

The proposed CNN model is successfully validated using simulated and experimental tests to check the classification accuracy. The proposed CNN model achieved average classification accuracies during the training and testing processes around 98.3% and 98.9% in simulation tests, and 96.76% and 97.41 % in experimental tests, respectively. It has also enormous efficiency during the rapid variation of the weather conditions of solar irradiance and temperature of the PV system. From the experimental tests observation, this high classification accuracies for the typical faults enhanced the model capability to be utilised in large-scale PV systems. In addition, it helps to avoid unpredicted power interruptions. In addition, the comparison with other traditional and AI models concludes that the proposed model provides automatic fault classification for both temporary and permanent faults with low cost of implementation.

The proposed CNN model can be extended for more practical fault classification to examine partial shading faults considering different patterns and irradiance levels, multi-fault case studies, and different PV configurations.

Nomenclature

$I_{m,mod}$	the module current at maximum power (A)
I_m	the array current at maximum power (A)
I_o	the diode current (A)
I_{ph}	photo-generated current (A)
$I_{sc,mod}$	the module short circuit current (A)
$I_{sc,ref}$	the short circuit current of reference module at S, and T (A)
$N_{p,mod}$	number of parallel strings = 3
N_p	number of parallel cells = 1
$N_{s,mod}$	number of series modules = 13
N_s	number of series cells = 60
P_{max}	the module t maximum power (W)
R_s	the equivalent series resistance of module (Ω)
R_{sh}	the shunt resistance of module (Ω)
S_{STC}	the solar irradiance at STC = 1000 (W/m^2)
T_{STC}	the surface temperature at STC = 25 (°C)
$V_{m,mod}$	the module voltage at maximum power (V)
V_m	the array voltage at maximum power (V)
$V_{oc,mod}$	the module open-circuit voltage (V)
$V_{oc,ref}$	the module open circuit voltage of reference module at S, and T (V)
V_{th}	the thermal voltage (V)

FC	false correct
FF	the fill factor
FN	false negative
I	the module output current (A)
K	Boltzmann's constant
S	the solar irradiance (W/m^2)
T	the module temperature (°C)
TC	true correct
V	the module output voltage (V)
n	the ideality factor
q	the charge of the electron(Col.)
α	the temperature coefficient of $I_{m,mod}$ (A/C)
β	the temperature coefficient of $V_{m,mod}$ (V/C)

ORCID

Sayed A. Zaki https://orcid.org/0000-0001-7545-0039
Ahmed Rabee Sayed https://orcid.org/0000-0001-5855-7125

REFERENCES

1. Mahmoud, Y.A., et al.: A parameterization approach for enhancing PV model accuracy. IEEE Trans. Ind. Electron. 60(12), 5708–5716 (2012)
2. Almonacid, F., et al.: Review of techniques based on artificial neural networks for the electrical characterization of concentrator photovoltaic technology. Renewable Sustainable Energy Rev. 75(75), 938–953 (2017)
3. Chen, Z., et al.: Intelligent fault diagnosis of photovoltaic arrays based on optimized kernel extreme learning machine and IV characteristics. Appl. Energy 204(204), 912–931 (2017)
4. Zaki, S.A., Zhu H., Yao J., Fault detection and diagnosis of photovoltaic system using fuzzy logic control. In: E3S Web of Conferences, EDP Sciences, 1–6 (2019)
5. Yu, K., et al.: Parameters identification of photovoltaic models using an improved java optimization algorithm. Energy Convers. Manage. 150, 742–753 (2017)
6. Oliva, D., et al.: Parameter estimation of photovoltaic cells using an improved chaotic whale optimization algorithm. Appl. Energy 200, 141–154 (2017)
7. Ma, J., et al.: Parameter estimation of photovoltaic models via cuckoo search. Eur. J. Appl. Math. (2), 1–11 (2013)
8. Chen, X., et al.: Teaching-learning-based artificial bee colony for solar photovoltaic parameter estimation. Appl. Energy 212(15), 78–88 (2018)
9. Oliva, D., et al.: Parameter identification of solar cells using artificial bee colony optimization. Energy 72(7), 93–102 (2014)
10. Li, S., et al.: Parameter extraction of photovoltaic models using an improved teaching-learning-based optimization. Energy Convers. Manage. 186(186), 293–305 (2019)
11. Zhao, Y., et al.: Graph-based semi-supervised learning for fault detection and classification in solar photovoltaic arrays. IEEE Trans. Power Electron. 30(5), 2848–2858 (2015)
12. Guerriero, P., et al.: Monitoring and diagnostics of PV plants by a wireless self-powered sensor for individual panels. IEEE J. Photovoltaics 6(1), 286–294 (2015)
13. Yi, Z., et al.: Line-to-line fault detection for photovoltaic arrays based on multi-resolution signal decomposition and two-stage support vector machine. IEEE Trans. Ind. Electron. 64(11), 8546–8556 (2017)
14. Chine, W., et al.: A novel fault diagnosis technique for photovoltaic systems based on artificial neural networks. Renewable Energy 90(90), 501–512 (2016)
15. Zhu, H., et al.: Fault diagnosis approach for photovoltaic arrays based on unsupervised sample clustering and probabilistic neural network model. Sol. Energy 176(176), 395–405 (2018)
16. Zhao, Q., et al.: A new PV array fault diagnosis method Using Fuzzy C-Mean clustering and Fuzzy membership algorithm. Energies 11(1), 238–247 (2018)

17. Lee, D., et al.: Convolutional neural net and bearing fault analysis. Proceedings of the International Conference on Data Mining (DMIN). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 194–200 (2016)

18. Chen, Z., et al.: Gearbox fault identification and classification with convolutional neural networks. Shock Vib. 2015, 1–10 (2015)

19. Li, X., et al.: Deep learning based module defect analysis for large-scale photovoltaic farms. IEEE Trans. Energy Convers. 34(1), 520–529 (2018)

20. Fadhel, S., et al.: "Data-driven approach for isolated PV shading fault diagnosis based on experimental IV curves analysis." In: 2018 IEEE International Conference on Industrial Technology (ICIT). IEEE, 927–932 (2018)

21. Zaki, S.A., et al.: "Detection and localization the open and short circuit faults in PV system: A MILP Approach." 2020 Asia Energy and Electrical Engineering Symposium (AEEES). IEEE, 187–193 (2020)

22. Satpathy, P.R., Sharma, R.: Power and mismatch losses mitigation by a fixed electrical reconfiguration technique for partially shaded photovoltaic arrays. Energy Convers. Manage. 192, 52–70 (2019)

23. Chander, S., et al.: Impact of temperature on performance of series and parallel connected mono-crystalline silicon solar cells. Energy Rep. 1, 175–180 (2017)

24. Alam, M.K., et al.: A comprehensive review of catastrophic faults in PV arrays: Types, detection, and mitigation techniques. IEEE J. Photovoltaics 5(3), 982–997 (2015)

25. Lyden, S., Haque, M.E.: Modelling, parameter estimation and assessment of partial shading conditions of photovoltaic modules. J. Mod. Power Syst. Clean Energy 7(1), 55–64 (2019)

26. Lee, K.B., et al.: A convolutional neural network for fault classification and diagnosis in semiconductor manufacturing processes. IEEE Trans. Semicond. Manuf. 30(2), 135–142 (2017)

27. Qu, J., et al.: Adaptive fault diagnosis algorithm for rolling bearings based on one-dimensional convolutional neural network. Chin. J. Sci. Instrum. 39, 134–143 (2018)

28. Dhimish, M., Holmes, V.: Fault detection algorithm for grid-connected photovoltaic plants. Sol. Energy 137, 236–245 (2016)

29. Ali, M.H., et al.: Real time fault detection in photovoltaic systems. Energy Procedia 111, 914–923 (2017)

30. Wang, Y., Li, J.: Credible intervals for precision and recall based on a K-fold cross-validated beta distribution. Neural Comput. 28(8), 1694–1722 (2016)