



Fault diagnosis for PV arrays considering dust impact based on transformed graphical features of characteristic curves and convolutional neural network with CBAM modules

Jiaqi Qu ^a, Qiang Sun ^b, Zheng Qian ^{a,*}, Lu Wei ^c, Hamidreza Zareipour ^d

^a School of Instrumentation and Optoelectronic Engineering, Beihang University, Beijing, China

^b State Key Laboratory of Control and Simulation of Power System and Generation Equipment, Department of Electrical Engineering, Tsinghua University, Beijing, China

^c School of Electronics and Information Engineering, Beihang University, Beijing, China

^d The Department of Electrical and Computer Engineering, University of Calgary, Calgary, Canada

HIGHLIGHTS

- The proposed technique provides universal ability to diagnose different PV configurations with dynamic array faults.
- The Isc-Voc normalized GADF transformation method can extract distinguishing features without calibrations.
- The CNN-CBAM model can accurately classify and identify the full range of dynamic complicated faults.
- The novel fault diagnosis method extends the application under the influence of dust in various scenarios.

ARTICLE INFO

Keywords:

Photovoltaic
Fault diagnosis
Dust impact
Graphical feature transformation
Characteristic curves

ABSTRACT

Various faults can occur during the operation of photovoltaic (PV) arrays, and both dust-affected operating conditions and various diode configurations complicate the faults. However, current methods for fault diagnosis based on I-V characteristic curves usually do not effectively use all the distinguishable information contained in the I-V curves or often rely on calibrating the field characteristic curves to standard test conditions (STC). It is difficult to apply these methods in practice and accurately identify multiple complex faults with similarities in different blocking diode configurations of PV arrays under the influence of dust. Therefore, a novel fault-diagnosis method for PV arrays that considers the impact of dust is proposed. In the pre-processing stage, the Isc-Voc normalized Gramian angular difference field (GADF) method is presented, which normalizes and transforms the resampled PV array characteristic curves from the field, including I-V and P-V, to obtain transformed graphical feature matrices. Subsequently, in the fault diagnosis stage, the convolutional neural network (CNN) model with convolutional block attention modules (CBAM) is designed to classify faults, which identifies complex fault types from the transformed graphical matrices containing complete discriminative fault information. In addition, the performances of different graphical feature transformation and CNN-based classification methods are compared using case studies. The results indicate that the developed method for PV arrays with different blocking diode configurations under various operating conditions has high fault diagnosis accuracy and reliability.

1. Introduction

Owing to its advantages of cleanliness, accessibility, and utilization, solar energy has been gaining increasing attention in the fight against global warming and fossil fuel energy shortages. The International Energy Agency (IEA) predicts that the global PV market will grow by 25%

year-on-year to 197 GW of newly installed capacity by 2022, with a cumulative installed capacity exceeding 1000 GW. From 2022 to 2031, The global PV grid-connected installed capacity is expected to increase at an average annual rate of 8% [1]. However, PV systems are often subjected to abnormal failures owing to disturbances in the operating environment, which can result in an estimated annual energy loss of up

* Corresponding author at: Xueyuan Road No.37, Haidian District, Beijing 100191, China.

E-mail address: qianzheng@buaa.edu.cn (Z. Qian).

to 18.9% [2].

PV arrays, which are the core components of PV systems, are prone to multiple faults such as short circuits, open circuits, abnormal degradation, and partial shading in complex outdoor environments [3]. Currently, traditional DC-side protection methods such as overcurrent protection (OCPDs) and ground fault protection (GFPDs) have certain difficulties in determining the type of fault [4,5]. Furthermore, owing to the nonlinear output characteristics of PV arrays, both low-mismatch faults and faults at low irradiances may cause protection failures [6]. These failures not only affect power generation, but also lead to serious safety issues. In certain cases, faults may increase the risk of fire and personal danger if not detected or corrected in time [7]. In addition, dusty environments have a significant impact on the operating performance of PV systems [8–10]. Dust from PV panels can reduce the power of PV systems [11]. More importantly, the long-term dust deposition operating conditions also complicate faults, forming compound faults that are more difficult to classify [12]. Therefore, the effective fault detection and diagnosis of PV arrays under the influence of dust is essential for the safety and reliability of PV systems [13].

Currently, there are two main categories of techniques for detecting and diagnosing faults in PV arrays. One is the offline diagnosis methods, which include earth capacitance measurement (ECM) [14], time-domain reflection (TDR) [15,16], and thermal methods (VTMs) [17]. These methods rely on specific external signal generators for testing and require the offline operation of the PV array, which interfere with normal operation and make it difficult to diagnose faults online. The other category is the online diagnostic methods that do not affect the normal operation of PV systems. According to the diagnostic indicators, these online methods are further divided into those based on the measurement of the current and voltage output on the DC side of the system [4,18,19], those based on the measurement of the current and voltage output on the AC side of the system [6,20,21], and those based on the I-V characteristic curves [22–24]. Specifically, some DC indicator-based methods rely on a large number of sensors installed in each substring, which is costly and has limited universality [25,26]. Other DC and AC indicator-based methods do not require the installation of a large number of sensors, but they exist some remaining issues. For example, diagnosis using residual thresholds or ratios of current and voltage output on the DC/AC side of the system to distinguish fault types is insufficient [27,28], and it is difficult to fully and accurately diagnose multiple complex faults [13]. Because I-V characteristic curves usually contain rich information on the status of PV modules, diagnosis based on I-V curves is a popular topic [29]. Moreover, the I-V tracer currently supports measurements for a single module or small-scale strings or arrays and has realized online measurements without changing the operational state [30,31]. In this sense, the diagnostic method based on I-V curves can be applied to all common PV installations and is easier to implement in the field.

In existing studies, according to the forms of curve features applied, fault diagnosis methods using I-V characteristic curves can generally be divided into i) raw I-V curves relying on deep learning models for diagnosis and ii) extracted key features of I-V curves for diagnosis. The former includes (1) taking the I-V curve data as input directly. For example, Chen et al. [32] assembled I-V curve with irradiance (G) and module temperature (T) into a 4-column matrix to classify eight classes of PV array faults using an improved ResNet model. Gao et al. [33] designed a fusion model of convolutional neural network (CNN) and residual-gated recurrent unit (Res-GRU) to diagnose hybrid faults using a combination matrix of I-V curves, irradiance, and temperature as inputs. (2) taking the residual between the measured I-V curve and the theoretical curve as the input. For example, Chine et al. [34] compared the difference between the measured and simulated PV array output powers and identified faults based on the attributes of the differences in I-V curves. Liu et al. [22] proposed a fault diagnosis method based on stacked autoencoder (SAE) and clustering, which extracts features from the difference between the simulated and measured I-V curves to

achieve classification. These methods do not perform additional feature processing on the I-V curves but rely on deep learning models to distinguish between different fault types, which often require classifiers with complex network structures [35]. In addition, their fault discrimination ability is limited owing to the absence of feature processing, even if they include all I-V variables, but do not possess the full fault discrimination information embedded in the variables. They only enable the classification of fixed-level faults and have a small diagnostic ability for dynamic variable parameter fault defects [29]. The latter includes (3) identifying the key features of array characteristics (e.g. V_{OC} , I_{SC} , V_{MPP} , I_{MPP} , FF , R_s , and R_p) in the curves as inputs. For example, Fadhel et al. [36] adopted V_{MPP} , I_{MPP} , and P_{MPP} as features to classify four shading configurations. Liu et al. [37] extracted five key points from I-V curves as valid features and inputted to a fault diagnosis method based on a variable prediction model. Besides, similar approaches were presented in [38–40]. (4) calculating the shape features (e.g. derivative and curvature) of the curves as inputs. For example, Bressan et al. [41] proposed a method based on the analysis of the first and second derivatives of the I-V curves for detecting faults in series resistors and the activation of bypass diodes. Ma et al. [42] analyzed the extraction of negative peaks on the derivative of the I-V curves, whereby single faults and compound faults at different levels of shading were diagnosed. However, it should be pointed out that these studies only used part information of characteristic curves. The diagnosis was completed by analysing the current (I_{MPP}), voltage (V_{MPP}), power (P_{MPP}), and curve-shaped features at the maximum power point (MPP). In some complex scenarios such as faults considering soiling, the feature extraction process is complicated, and they may appear to have the same MPPs, leading to incorrect classifications.

Several recent studies have made additional contributions to this field. Lin et al. [43] extracted multi-scale fault features using different scales of convolution horizons and identified fixed fault parameters of faults affected by soiling. Huang et al. [44] investigated full-scale faults under operating conditions with soiling impact in PV arrays without blocking diodes, and Li et al. [35] used full characteristic information of the graphical features of I-V curves and machine learning techniques for PV array fault diagnosis. However, several gaps remain in the literature. First, the development of faults is dynamic and can occur at any stage during diagnosis (i.e. the parameters that characterize the type of fault are variable). For practical engineering applications and reasonable diagnosis, faults of the same type at various stages of different severities (i.e. full-scale faults) should be identified as the same type. To the best of our knowledge, most existing fault diagnosis methods focus only on fixed-level faults. No studies have provided a comprehensive analysis of full-scale faults with variable fault parameters in PV arrays with various blocking diode configurations considering the impact of soiling. Second, extracting the features of array characteristics or curve shapes does not make effective use of the full information contained in the I-V curves, leaving incomplete identification of PV array faults. Furthermore, most current diagnostic methods using characteristic curves rely on additional measured curves of fault samples to fit the relationship models of the key parameters or to calibrate the I-V curves to the STC conditions [44,45], severely restricting their practical applicability. Therefore, this study aims to develop a feature processing method using the full information of the characteristic curves that does not require additional experiments to calibrate the curves under different environments. Furthermore, a full-range multifault classification method adapted to different PV array configurations considering the soiling impact is designed. This enables the complete diagnosis of complex faults under various operating conditions and PV array configurations. The main contributions of the proposed method are as follows:

- (1) Various fault types of PV arrays in multiple scenarios are compared and analyzed, including PV arrays with and without blocking diode configurations, and compound fault types with and without the influence of soiling. This overcomes the

challenge of the uniqueness of array fault diagnosis methods for different objects and provides a universal ability to diagnose different PV configurations in any state during the dynamic evolutionary period of variable parameter faults.

- (2) A graphical feature transformation method based on the I_{sc} - V_{oc} normalized GADF is proposed to extract common features of the same fault in different environments, and the transformed graphical features of the characteristic curves are stacked into 2-channel feature matrices as input features for the classification model. This fills the gap in diagnosing complex faults affected by soiling using fully distinguishable information contained in the characteristic curves without calibration experiments, which greatly improves its practical application value.
- (3) A classification model of CNN with CBAM modules is designed for fault diagnosis in multiple scenarios that can accurately classify and identify the full range of complicated faults. This extends the performance of diagnosis methods under the influence of dust and improves diagnosis accuracy and robustness under multiple scenarios.

The rest of this paper is organized as follows: Section 2 formulates the preliminary analysis of the problem to be studied. Section 3 details the proposed fault diagnosis methodology. Section 4 presents experimental results and discussion. Finally, Section 5 concludes this work.

2. Preliminary analysis

In this section, first, the fault behaviors of PV arrays with different blocking diode configurations are analyzed. Second, the characteristic curves under normal operating condition and condition considering the influence of soiling are compared. Furthermore, the pre-processing techniques used for current fault diagnosis are investigated, and the limitations of the existing pre-processing methods are identified.

2.1. Faults analysis of PV arrays

As shown in Fig. 1a two configurations (with and without blocking diodes, denoted as Configuration 1 and 2, respectively) are designed to analyze each type of fault. Various faults can occur during the operation of a PV array. These include short circuits in substrings, open circuits of substrings, various degrees of shading, series resistance degradation of the array, and parallel resistance degradation of the array.

PV arrays with different blocking diode configurations differ in the manifestation of short circuits [6], as exemplified by the I-V characteristic curves shown in Fig. 2. Specifically, the open-circuit voltage of a normal substring (denoted V_{OCS}) of a PV array is proportional to the number of serial modules (denoted N_{SM}), and when different numbers of PV modules in a particular substring are short-circuited (denoted N_{LL}),

then the OC voltage of the fault string is $(N_{SM} - N_{LL})/N_{SM} V_{OCS}$ (denoted as V_{OCF}). When the substrings are equipped with blocking diodes, that is, Configuration 1, the current of each string is allowed to flow only in one direction. The V_{OCS} and V_{MPP} of the faulty string decrease, whereas I_{MPP} remains unchanged. When $V_{OCS} > V_{OCF}$, the faulty string is disconnected from the array, which is expressed in the IV/PV characteristic curve as the presence of a local minimum inflection point. The overall V_{OC} of the array remains stable, as shown in Fig. 2b. In contrast, when the substrings are not equipped with blocking diodes, that is, Configuration 2, the current from the normal string reverses into a faulty string when $V_{OCS} > V_{OCF}$. Consequently, the V_{OC} of the faulty array with short circuits is lower than that of the normal array. Notably, when $V_{OCS} < V_{OCF}$, the characteristic curves of the PV array are the same for both configurations.

In addition, during outdoor operation, all modules in the PV array can be affected by various types of shading, including shading from buildings, tree shade, etc., as well as shielding from dust adhering to the surface. The essential effect of both is manifested as a reduction in the irradiance intensity of the affected modules. Specifically, the former generally results in a larger reduction in irradiance, typically 20% or more, whereas dust accumulation is considered a specific form of shading that occurs across all the PV modules of an array and produces a relatively small reduction in irradiance, usually below 20% [46,47].

The enlarged structure of the PV module in Fig. 1a shows that every N cell is inversely connected in parallel to one bypass diode. For a single module, when the shading intensity and area are fixed, the output I-V characteristic curve of the module exhibits a single peak, as shown in Fig. 3a. Furthermore, when a module in a substring containing multiple modules is shaded, the I-V curve splits into two parts [36]. When the voltage is lower than the substring voltage (denoted V_S), the shaded cells of the shaded module are bypassed by the bypass diode, and the I-V characteristics in this interval are approximately equivalent to those of the remaining unshaded modules in series. When the voltage exceeds V_S , the bypass diode is turned off, resulting in a decrease in the total output current of the string [48]. The I-V curve in this interval is mainly determined by the shaded module. Therefore, regardless of the degree of shading, there is a turning peak in the I-V/P-V characteristic curve. The degree of the shadow can influence the inflection point of the current, as shown in Fig. 3a. The current at the inflection point decreases with increasing shading degree and number of shading modules. As the degree of shading increases, the voltage at the inflection point proportionally decreases, as shown in Fig. 3b. It also can be seen that in Fig. 2 and Fig. 3b, the characteristic curves of arrays under certain shading conditions are very similar to the short-circuited arrays configured with blocking diodes.

In fact, the intensity of dust accumulation is uneven across all modules, representing a non-uniform soiling effect on the array [44,46], as shown in Fig. 1b. Therefore, under the influence of non-uniform

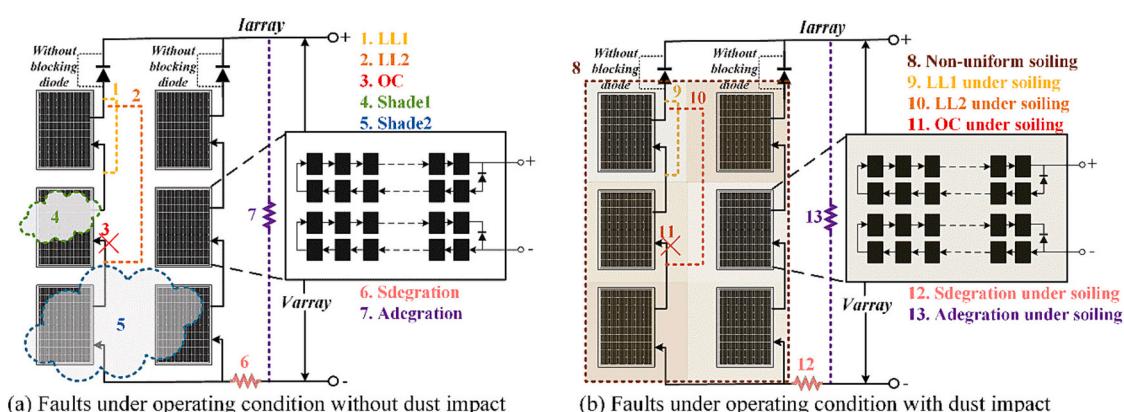


Fig. 1. Faults of PV arrays under different operating conditions (a schematic diagram for understanding, not a closed-loop circuit).

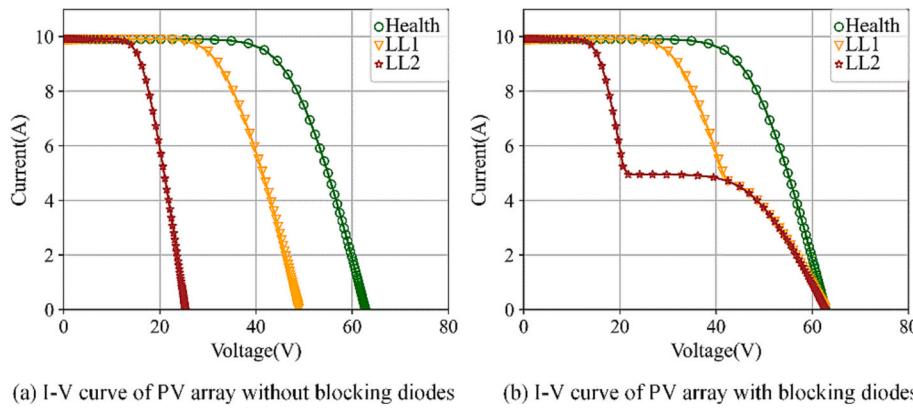


Fig. 2. I-V curve of PV arrays with different blocking diode configurations.

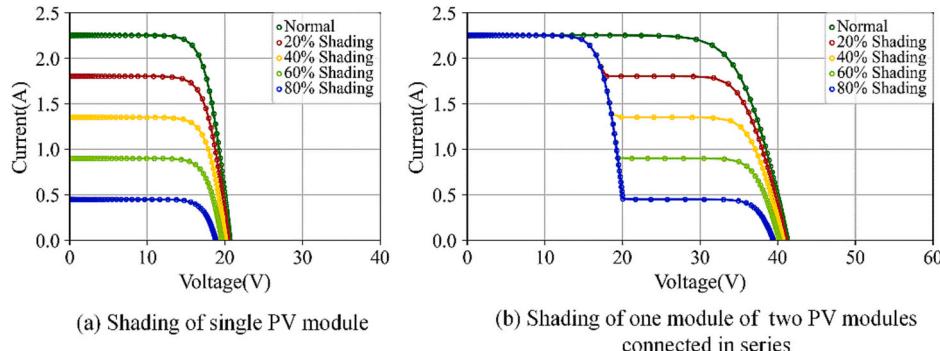


Fig. 3. I-V curve of PV modules with different structures under the influence of different shading levels.

soiling, each module of the array is equivalent to being affected by different degrees of shading, and the characteristic curves of the array also show a different number of peaks. Importantly, when other faults occur under the impact of dust, such as short and open circuits, the characteristic curves become more complex, showing a superposition of multiple peaks and the corresponding fault characteristics simultaneously, as shown in Fig. 4b. However, most cleaning operations for dust or soiling are usually carried out annually or more frequently in areas heavily affected by dust, but other faults under the influence of non-uniform soiling can still occur during the cleaning interval [49,50]. This makes the diagnosis of complex fault types more challenging and difficult to accurately identify.

It can be seen that various complex fault types of different array structures are rich and diverse and have high similarities. Therefore, it is

important to develop an effective method for fault diagnosis under multiple scenarios.

2.2. Pre-processing methods for fault diagnosis

The shape of the PV characteristic curves depends on environmental conditions, such as irradiance and temperature [51]. Therefore, when using characteristic curves for PV array fault diagnosis, the influence of different environmental conditions causing different characteristic manifestations of the same fault type should be excluded to reduce the feature confusion caused by factors other than the fault types [35]. At present, there are three data pre-processing methods for minimizing the impact of environmental factors on the characteristic curves:

(1) The input feature is a two-dimensional matrix recombined by I-V

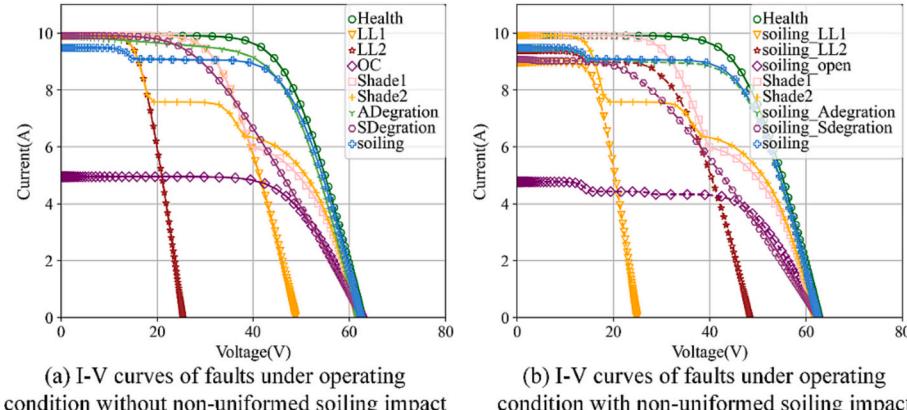


Fig. 4. I-V curves of faults under different operating conditions.

curve and ambient variables, which stems from the ability of the convolutional neural network to automatically extract the features of two-dimensional data. Specifically, irradiance and temperature are repeatedly supplemented into column vectors with the same points as the I/V vector and then stitched together with the I-V curve to form a final two-dimensional matrix, as shown in Fig. 5, which is used as input for the deep-learning-based classification methods.

This method does not directly eliminate the influence of environmental variables on the characteristic curves but relies only on the convolutional network to extract common feature information of the same fault type under different environmental conditions. It can be inferred that its ability to handle complicated fault diagnoses is shallow.

(2) Converting key features of array characteristics in different environments to features under STC is often applied to diagnostic methods using key features identified in characteristic curves. That is, the open-circuit voltage V_{OC} , short-circuit current I_{SC} , maximum power point voltage V_{MPP} , maximum power point current I_{MPP} , and equivalent series resistance R_S are converted to the corresponding values under STC. The approach for obtaining the feature functions is based on traditional approximation equations [52,53], where the unknown parameters are denoted by a , b , c , and d . The functions of the key features can be rewritten as:

$$V_{OC} = V_{OC,STC} + a_1 \cdot \ln \frac{G}{G_{STC}} + a_2 \cdot dT + a_3 \cdot \frac{G}{G_{STC}} dT \quad (1)$$

$$V_m = V_{m,STC} + b_1 \cdot \ln \frac{G}{G_{STC}} + b_2 \cdot dT + b_3 \cdot \frac{G}{G_{STC}} dT \quad (2)$$

$$I_m = c_1 \cdot I_{m,STC} \frac{G}{G_{STC}} + c_2 \cdot dT + c_3 \cdot \frac{G}{G_{STC}} dT \quad (3)$$

$$R_S = R_{S,STC} \left(\frac{G}{G_{STC}} \right)^{d_1} + d_2 \cdot dT + d_3 \cdot \frac{G}{G_{STC}} dT \quad (4)$$

The parameters of this nonlinear static model must be identified by searching for multiple characteristic curves. In addition, as the curves under STC still behave differently for various fault types, separate parameter identifications are required for various faults with STC conditions, which involve additional test experiments with fault samples. Moreover, diagnostic methods that extract a limited number of metrics as features are ineffective in classifying complex fault types.

(3) Correcting all points of the entire I-V characteristic curve is a pre-processing method that relies on the use of the full I-V information for diagnosis. IEC 60891 [54] defines three standard procedures for the correction of I-V curves, which are used to compare curves measured

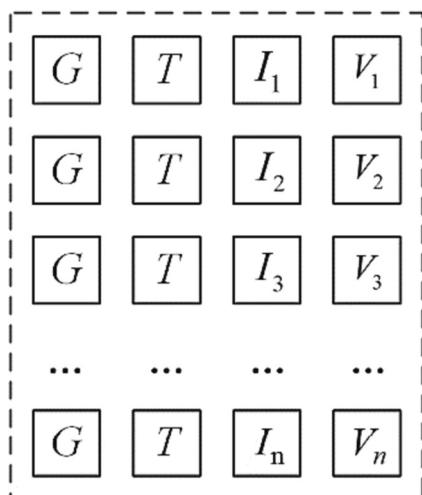


Fig. 5. The recombined GTIV matrix.

under different conditions, thus enabling health monitoring of PV panels. The following are named M1, M2, and M3, along with the correction method of M2new proposed in [45]:

M1:

$$I_2 = I_1 + I_{SC} \cdot \left(\frac{G_2}{G_1} - 1 \right) + \alpha \cdot (T_2 - T_1) \quad (5)$$

$$V_2 = V_1 - R_S \cdot (I_2 - I_1) - \kappa \cdot I_2 \cdot (T_2 - T_1) + \beta \cdot (T_2 - T_1) \quad (6)$$

M2:

$$I_2 = I_1 \cdot (1 + \alpha_{rel} \cdot (T_2 - T_1)) \cdot \frac{G_2}{G_1} \quad (7)$$

$$V_2 = V_1 + V_{OC1} \cdot \left(\beta_{rel} \cdot (T_2 - T_1) + a \cdot \ln \left(\frac{G_2}{G_1} \right) \right) - R_S \cdot (I_2 - I_1) - \kappa \cdot I_2 \cdot (T_2 - T_1) \quad (8)$$

M2new:

It uses the same equation as M2 for current correction, but corrects for voltage by replacing the term “ V_{OC1} ” in (8) with “ $V_{OC1} \cdot [1 + \beta_{rel}(25 - T_1)]$ ”.

M3:

$$I_3 = I_1 + \gamma(I_2 - I_1) \quad (9)$$

$$V_3 = V_1 + \gamma(V_2 - V_1) \quad (10)$$

$$G_3 = G_1 + \gamma(G_2 - G_1) \quad (11)$$

$$T_3 = T_1 + \gamma(T_2 - T_1) \quad (12)$$

In fact, M1, M2, and NewM2 are correction methods based on a single curve, which requires the setting of corresponding correction factors, such as κ and R_S . That is, after the correction parameters are determined, they can be directly corrected from any test condition to the STC. However, it is still difficult to determine the correction factors for PV panels onsite because of the rigorous experimental conditions required for the IEC 60891 procedure. In addition, research has shown that, owing to the differences in irradiance, module temperature, and severity of faults, these methods introduce significant errors, making it difficult to perform well under all fault conditions [45]. Moreover, the distortion of the curve shape in the IEC method typically leads to a relative error of 13.8%, and an estimation error of the fault features extracted from the correction curve also occurs frequently. If these features are used as defect features, it may affect the diagnosis of faults.

Alternatively, M3 does not contain correction factors, but the interpolation constant γ must be set. This method applies a linear interpolation method to multiple I-V characteristic testing curves to obtain the I-V characteristics under STC. Although the multiple-curve-based method (M3) generally offers better performance than the above methods based on a single curve, the conversion of a certain curve with both G and T requirements relies on multiple testing curves, and the measurement and calculation are relatively complex and inefficient. Therefore, they are not suitable for rapid field diagnosis.

The use of characteristic curves for PV system health monitoring and fault diagnosis is promising. Considering the limitations of the practical application of the pre-processing methods discussed above, it is beneficial to explore solutions based on field measurement data while reducing the dependence on the calibration process.

3. The proposed fault diagnosis methodology

In this section, the accurate modeling approach of a real PV array is presented, and the simulation method of faults in different configurations of PV arrays under different operating conditions is described. Subsequently, a data processing method that does not depend on the calibration of the characteristic curves to the STC is proposed to obtain

transformed graphical feature matrices with full information on the characteristic curves. Finally, a fault diagnosis model based on the convolutional neural network with CBAM modules is introduced.

3.1. Configuration of the simulated PV modeling

To explore the complex faults of PV arrays under different operating conditions, the single-diode PV cell model proposed in [55] is used, as shown in Fig. 6. Based on this model, a PV module model is built on the PSCAD/EMTDC platform with configuration parameters from the Shell Solar SP-70 datasheet listed in Table 1. We compared the simulated I-V curves of this PV module model under different radiation intensities at 25 °C with the manual curves supplied by the manufacturer. There is a high correlation between the two, which verifies the consistency of equivalence between the model and the actual PV module. Based on this PV module model, the PV array model used in this study is further designed for a structure with three modules in series and two substrings in parallel. Among them, substrings containing blocking diodes result in two configuration types of PV arrays, whose structures are shown in Fig. 1. To characterize the faults under the operating environmental conditions of a real PV array as much as possible, we use one year of outdoor irradiance and temperature records measured from the PV plant as environmental control variables for the PV array model.

To demonstrate the generalizability of the proposed fault diagnosis method, the faults of PV arrays with two types of blocking diode configurations under different operating conditions are validated, including contamination-prone and ideal normal operating conditions. The two operating conditions comprised 14 and 9 types of faults, respectively, as follows:

Case 1. PV arrays under operating condition with non-uniform soiling impact.

- 1) Two types of line-to-line short circuits (LL): one or two modules in one string are shorted (noted as LL1 and LL2 respectively).
- 2) Open circuits (OC): one string is open.
- 3) Two types of shading (Shade): 1 or 2 modules are shaded to different degrees (noted as Shade1 and Shade2, respectively).
- 4) Series resistance degradation of the array (Sdegradation): an increase in the equivalent series resistance of array.
- 5) Parallel resistance degradation of the array (Adegradation): a decrease in the equivalent parallel resistance of array.
- 6) Non-uniform soiling (Soiling): the accumulation of soiling with varying degrees on the surface of each PV module in the presence of contamination.
- 7) Line-to-line short-circuit under the impact of non-uniform soiling (soiling_LL): 1 or 2 modules in one string are shorted under

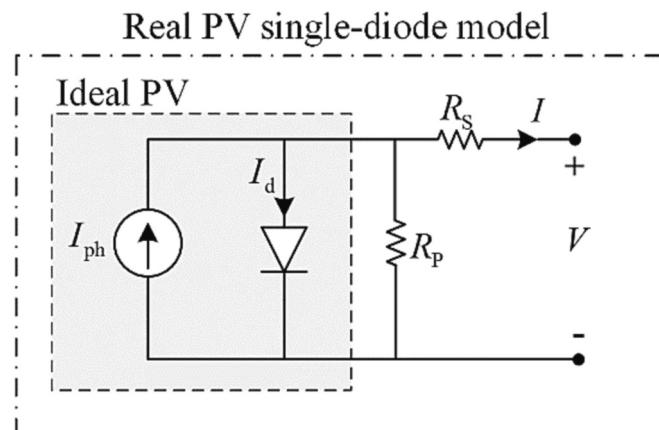


Fig. 6. The single-diode PV cell model.

Table 1

Electrical characteristics of solar module Shell SP-70 at nominal condition (25 °C and 1000 W/m²).

Parameter	Value
I_{SC} (A)	4.7
V_{oc} (V)	21.4
I_{MPP} (A)	4.25
V_{MPP} (V)	16.5
K_V (mV/°C)	-76
K_i (mA/°C)	2
N_S	36
R_s (Ω)	0.41
R_p (Ω)	141

varying soiling accumulation (denoted as soiling_LL1 and soiling_LL2, respectively).

- 8) Open-circuit under the impact of non-uniform soiling (soiling_OC): one string is open under varying soiling accumulation.
- 9) Series resistance degradation of the array under the impact of non-uniform soiling (soiling_Sdegartion): an increase in the equivalent series resistance of array under varying soiling accumulation.
- 10) Parallel resistance degradation of the array under the impact of non-uniform soiling (soiling_Adegartion): a decrease in the equivalent parallel resistance of array under varying soiling accumulation.

Case 2. PV arrays under operating condition without non-uniform soiling impact.

The types of faults include 1) - 6) above under the ideal operating condition without the influence of non-uniform soiling.

As an example, the characteristic curves for faults of the PV array with the two blocking diode configurations in Case 1, including I-V and P-V are shown in Fig. 7 and Fig. 8.

Importantly, what is different from [19,22,43] is that the simulations of faults in this study for Shade, Adegradation, SDegradation, and Soiling, as well as the severity of these faults, are not simply set as constants. This is because these faults change dynamically with time in service; therefore, the simulations should contain full-scale faults with variable fault parameters. In this study, the environmental conditions of a real PV plant are used as inputs for the temperature and irradiance of the PV array model, and the fault parameter of each fault type is set randomly within its corresponding full-scale parameter range to obtain characteristic curves. Specifically, for shading, the irradiance gain is set from 20% (low shading) to 100% (full shading) for one or two modules; for soiling, a special form of shading, the irradiance gain is set randomly within 10% for all modules to simulate non-uniform soiling accumulation; for abnormal aging of Adegradation and Sdegradation, the aging resistance values are set randomly for degradation in [20 Ω, 200 Ω] and [1 Ω, 15 Ω], respectively, adjusting for different levels of fault severity. Considering the dust-influenced operating conditions, faults are simulated by superimposing other faults alongside non-uniform soiling, where the degree of soiling and variable fault parameters are set randomly for dynamically growing fault types. The representation of different faults with full-scale variable fault parameters is shown in Fig. 9, where the characteristic curves show different distortion shapes.

3.2. Graphical feature transformation method of Isc-Voc normalized GADF

Outliers and noise in the collected characteristic curves can be eliminated in the data-cleaning phase using the method described in [56,57]. The characteristic curves with the anomalies removed are then

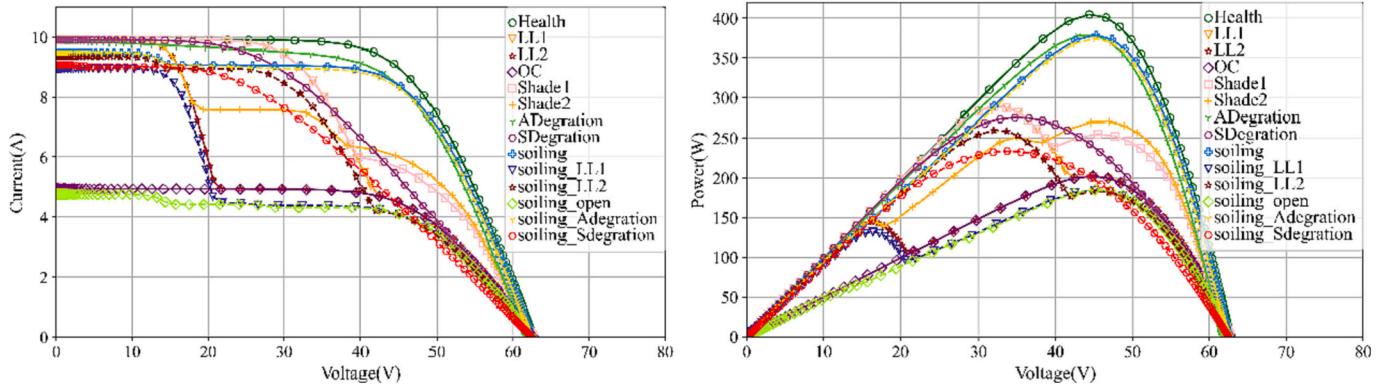


Fig. 7. Faults characteristic curves of PV array with blocking diodes.

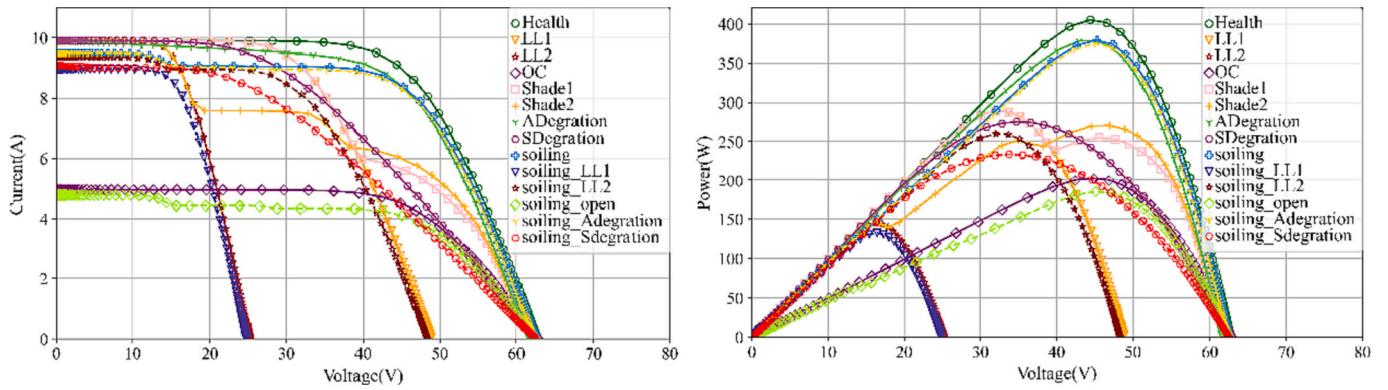


Fig. 8. Faults characteristic curves of PV array without blocking diodes.

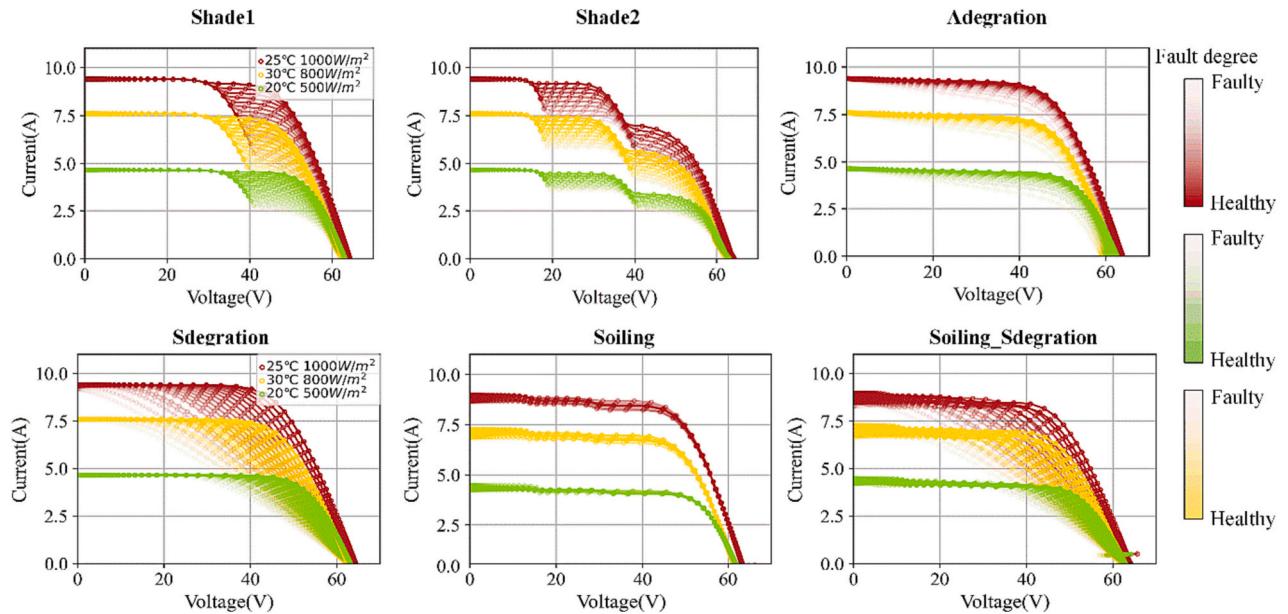


Fig. 9. I-V curves of different full-scale faults with variable fault parameters.

transformed into graphical features. In addition, we analyze in the Appendix the effect of directly processing raw noise-containing I-V curves for fault diagnosis using the following graphical feature transformation method. The graphical feature transformation method using the Gramian angular difference field (GADF) [58] enables the extraction of complete distinguishable information from the resampled characteristic

curves. Typically, GADF is applied to the transformation of time-series signals, which includes the following steps.

Step 1: Scale the data to [0,1] according to the following equation.

$$\tilde{x}_i = \frac{x_i - \min(X)}{\max(X) - \min(X)} \quad (13)$$

where the time series data is $X = \{x_1, x_2, \dots, x_N\}$ and its value at each timestamp is x_i , \tilde{x}_i is noted as the normalized value.

Step 2: Convert the rescaled sequence to the polar coordinate system, i.e. the values of the time series are treated as the cosine of the angle. The formulas for converting to polar coordinates are:

$$\begin{cases} \phi_i = \arccos(\tilde{x}_i), 0 \leq \tilde{x}_i \leq 1, \tilde{x}_i \in \tilde{X} \\ r_i = \frac{t_i}{N}, t_i \in N \end{cases} \quad (14)$$

where t_i is the time stamp of point x_i , N is the number of all the time points contained in the time series and \tilde{X} represents the rescaled time series.

Step 3: Obtain the angle difference of each pair, and then take the sine value for difference to form the GADF matrix:

$$GADF = [\sin(\phi_i - \phi_j)] = \sqrt{I - \tilde{X}^2} \cdot \tilde{X} \cdot \sqrt{I - \tilde{X}^2} \quad (15)$$

where I is the unit row vector $[1, 1, \dots, 1]$, r_i is the value of the polar axis, and ϕ_i the value of the polar angle.

When applied to a time series, the scaling of GADF uses full-time-axis data; thus, when applied to the scaling of the I-V/P-V characteristic curves, the V-axis should be analogous to the time-axis. If the GADF with normal normalization following the steps described above is used to convert the characteristic curves, that is, the conversion areas are A, B, C, and D in Fig. 10. The current V_{OC} and I_{SC} of each fault state are taken as their respective normalized maximum values, which would blur the differences in V_{OC} and I_{SC} between the different faults. We can observe from the GADF transformation results in Fig. 11 that when the GADF transformation is performed on the I- or P-axis with the normal normalization strategy, the same fault types under different irradiance and temperature conditions are relatively consistent in the shape of the converted graphical features, which can avoid misidentification of the same faults owing to different environments. However, similarities in shapes also exist between the characteristic curves of certain fault types such as health, LL1, LL2, and OC, as shown in Fig. 11. These curves differ in absolute values of V_{OC} and I_{SC} . Therefore, it is necessary to propose a new GADF transformation method that retains consistency in the characteristics of the transformation features for the same fault under different environmental conditions while having the ability to distinguish the similarity in the shapes of the characteristic curves for different fault types.

Through analysis of the single-diode equivalent model of the PV module in Fig. 6, the calculation equations for the short-circuit current

I_{SC} and open-circuit voltage V_{OC} are obtained as follows:

When the external circuit is short-circuited, that is, the load is zero, the short-circuit current I_{SC} is

$$I_{SC} = I_{ph} - I_0 \left[e^{\frac{I_{SC} R_S}{nV_T}} - 1 \right] - \frac{I_{SC} R_S}{R_p} \quad (16)$$

The dark current $I_d = I_0 \left[e^{\frac{I_{SC} R_S}{nV_T}} - 1 \right]$ flowing through the diode is very small and can be ignored. Considering that $R_p > R_S$, $\frac{I_{SC} R_S}{R_p}$ can also be ignored [53], then:

$$I_{SC} \approx I_{ph} \quad (17)$$

$$I_{ph} = [I_{SC,n} + k_i \cdot (T - 298)] \cdot \frac{G}{1000} \quad (18)$$

where T is the temperature in Kelvin, G is the irradiance, and k_i is the temperature coefficient of short-circuit current.

When the external circuit is open, i.e. when the load is close to ∞ and $I = 0$, the open-circuit voltage V_{OC} is [59]:

$$I = I_{ph} - I_0 \left[e^{\frac{I_{SC} R_S}{nV_T}} - 1 \right] - \frac{I_{SC} R_S}{R_p} \quad (19)$$

$$0 = I_{ph} - I_0 \left[e^{\frac{V_{OC}}{nV_T}} - 1 \right] - \frac{V_{OC}}{R_p} \quad (20)$$

$$V_{OC} = \frac{n k T}{q} \ln \left(\frac{I_{ph}}{I_0} + 1 \right) \quad (21)$$

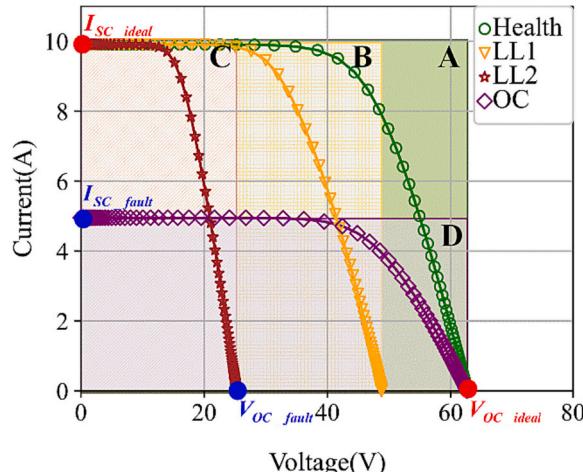
where the diode saturation current I_0 is expressed as [60,61]:

$$I_0 = I_{0,n} \left(\frac{T_n}{T} \right)^3 \exp \left[\frac{q E_g}{n k} \left(\frac{1}{T_n} - \frac{1}{T} \right) \right] \quad (22)$$

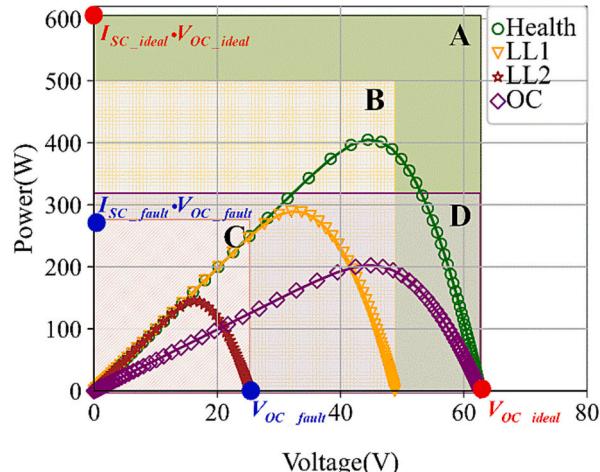
$$I_{0,n} = \frac{I_{SC,n}}{\exp(V_{OC,n}/nV_{t,n}) - 1} \quad (23)$$

$$V_{t,n} = \frac{N_s k T_n}{q} \quad (24)$$

where the band gap energy of the semiconductor is represented by E_g and the nominal saturation current at the nominal temperature T_n , i.e. 25 °C, is represented by $I_{0,n}$ via (23). The nominal thermal voltage represented by $V_{t,n}$ is shown in (24), $I_{SC,n}$ is the short-circuit current at nominal temperature, and $V_{OC,n}$ is the open-circuit voltage at nominal temperature. In addition, n is the ideal factor of the diode, N_s is the



(a) Normal normalization of similar faults in I-V curves



(b) Normal normalization of similar faults in P-V curves

Fig. 10. Normal normalization area of similar faults under the same ambient condition.

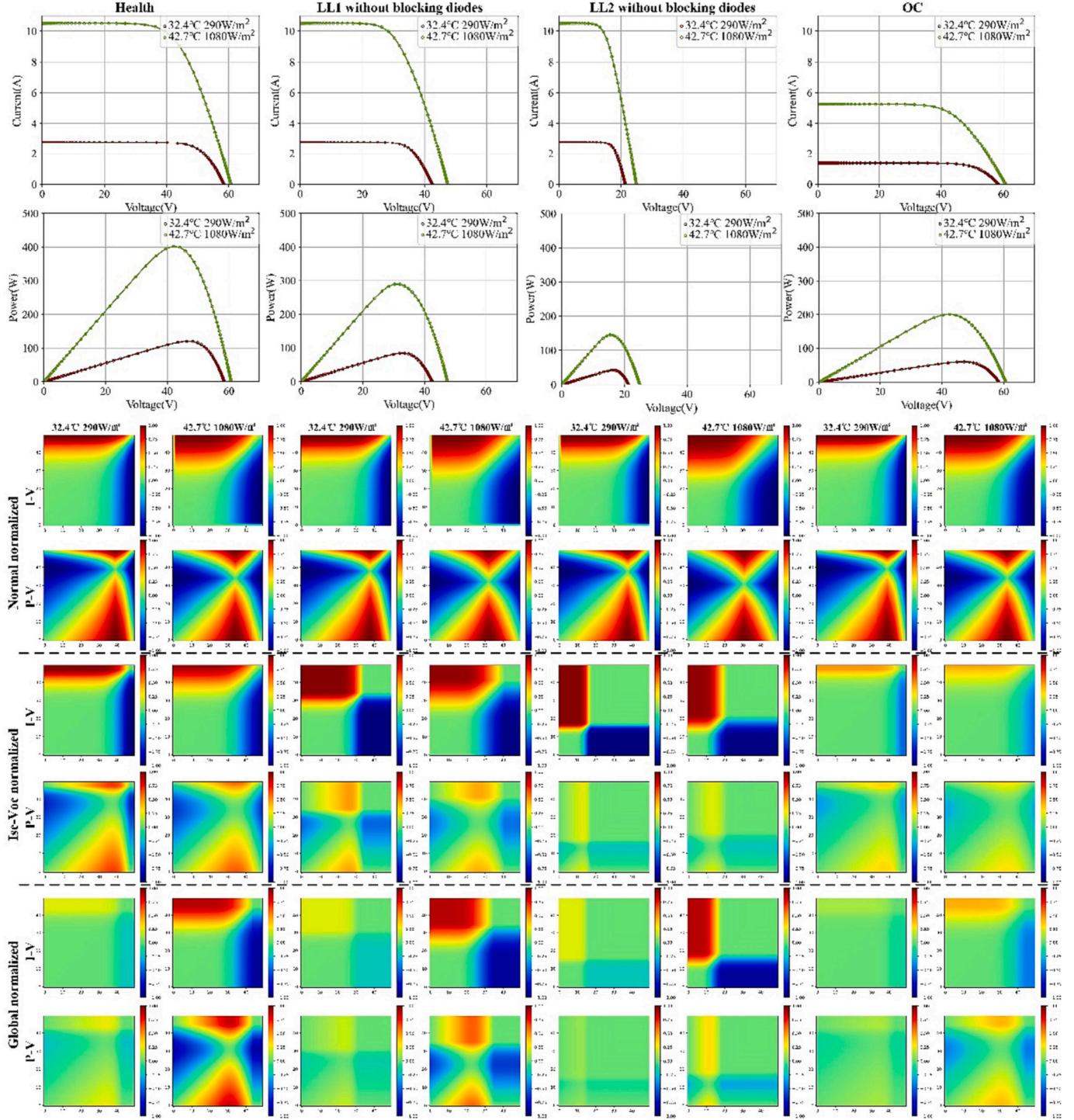


Fig. 11. Graphical matrices transformed by GADF with different normalization strategies for similar faults under different environmental conditions (the matrices are colored for only visualization).

number of series cells forming the module, k is the Boltzmann constant, and q is the charge quality. Thus, it can be seen from (22) that I_0 is only related to the temperature T .

It can be further concluded from (18) and (21) that the short-circuit current I_{SC} and open-circuit voltage V_{OC} of the PV modules depend only on the irradiance G and ambient temperature T , except for the constant factors with fixed values. This means that the ideal I_{SC} and V_{OC} are the same for different fault types under the same environmental conditions. Therefore, we propose a GADF method based on the normalization of I_{SC} and V_{OC} under environmental conditions, which can ensure that the

transformed graphical features of the same fault types are the same in different environments while having distinguishability between different faults with similar forms. Specifically, the main implementation process of the Isc-Voc normalized GADF method is as follows:

- (1) Calculate the ideal I_{SC} and V_{OC} (denoted as $I_{SC,ideal}$ and $V_{OC,ideal}$) of the ambient conditions corresponding to each characteristic curve, expand the maximum value of the V-axis to the ideal V_{OC} , and complement the current and power of the characteristic

- curve by zero over the range of the measured V_{OC} and the ideal V_{OC} .
- (2) Resample the complemented characteristic curves using the bilinear difference method to obtain uniformly distributed characteristic curves with small amounts of data. Specifically, the V axes of the I-V and P-V curves are within the range of [0, V_{OC_ideal}] at a uniform voltage interval, and the I- and P-axes are resampled at the same uniform interval, reducing the data of each curve from the original 200 points to 50 points.
 - (3) Convert the resampled characteristic curves to graphical features according to the Isc-Voc normalized GADF. That is, the V-, I-, and P-axes are normalized according to the ranges of [0, V_{OC_ideal}], [0, I_{SC_ideal}], and [0, $V_{OC_ideal} \times I_{SC_ideal}$], respectively.
 - (4) Calculate the inner product according to the differences in normalized angles, preserve the time dependence of the V-axis, and generate a GADF matrix of size 50^2 . The I-V or P-V characteristic curve corresponding to each environmental condition is transformed into a matrix, and the transformed I-V and P-V transformed matrices are stacked to form 2-channel feature matrices as inputs to the fault diagnosis model.

This method enables the selection of a unified V_{OC} under the same environmental conditions, which corresponds to the V-axis with the same timescale. Thus, all changes in the characteristic curves are reflected in the transformation matrices. The diagonal regions of the matrix shrank in different directions when the slope of the characteristic curve changed. Furthermore, the normalization strategy of selecting the maximum values of V_{OC} and I_{SC} in all characteristic curves without distinguishing environmental conditions is also compared and referred to as the global normalized GADF. A more detailed discussion and analysis of the three current universally applied methods (direct I-V, RP, and GADF) as well as the proposed normalization strategies are provided in Chapter 4.

3.3. Classification model of convolutional neural network with CBAM module

To improve the ability to diagnose complicated faults, we design a PV array fault diagnosis model of convolutional neural network with CBAM modules, referred to as CNN-CBAM, using the transformed

graphical feature matrices containing the full information of the characteristic curves as the input features.

The structure of the model is shown in Fig. 12 and is mainly composed of Convolution modules and CBAM attention modules.

The Convolution module can reduce the time and space dimensions and the number of free parameters required for training [62] owing to the benefits of the local receptive field and weight sharing. Therefore, the performance can be enhanced. Specifically, the Conv2d layer slides each filter in the input feature matrix through the local receptive field, calculates the sum of the dot products in the local field, and automatically extracts effective features from the inputs. The Pooling layer then divides the input area and calculates the average value of each area to complete the downsampling of the feature matrix.

The CBAM module focuses on important features and suppresses unimportant ones in the network, effectively improving the performance of CNN. The CBAM is composed of the channel attention module (CAM) and the spatial attention module (SAM) [63]. The detailed structures are shown in Figs. 13 and 14. Among them, the CAM emphasizes that the network should concentrate on useful channel features while ignoring other aspects using the maximum and average pools to compress the spatial dimension of the feature matrix. The SAM highlights that the network should focus on the local area of interest by applying the average and maximum pools along the channel dimension to retain the background information of the feature matrix. First, the input feature F is multiplied by the feature matrix $M_C(F)$ generated by CAM compression along the spatial dimension to obtain F' . The SAM then compresses F' along the channel dimension to generate the spatial feature matrix $M_S(F')$. Finally, the optimized feature matrix F'' is obtained by multiplying $M_S(F')$ and F' . The entire process can be represented as:

$$F' = M_C(F) \otimes F \quad (25)$$

$$F'' = M_S(F') \otimes F' \quad (26)$$

where \otimes represents multiplication between elements.

The CAM module processes and aggregates the input features with the maximum and average pools and then inputs them into a weight-sharing multilayer perceptron (MLP) network for summation, which is activated by the sigmoid function to generate the final channel attention feature $M_C(F)$. The calculation program for the CAM can be expressed as:

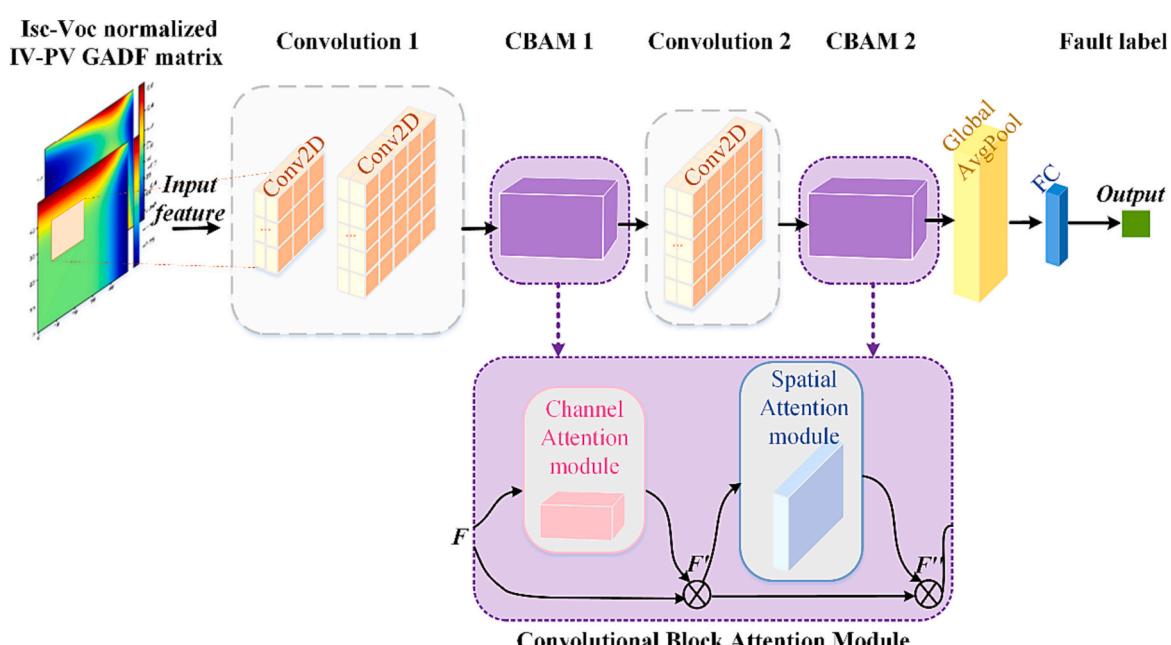


Fig. 12. The proposed classification model of CNN with CBAM modules.

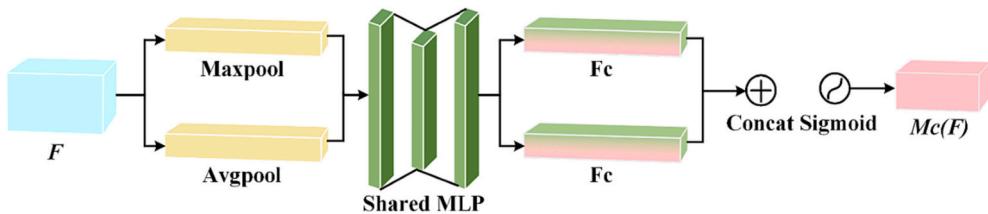


Fig. 13. The structure of channel attention module (CAM).

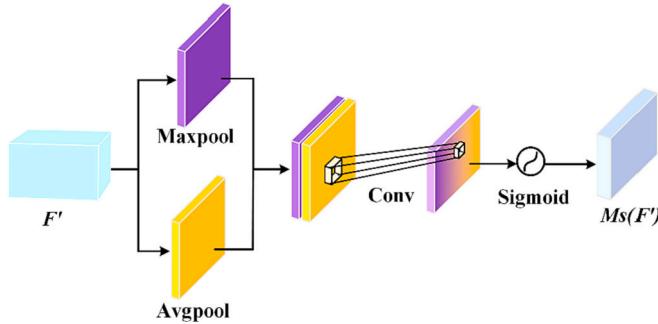


Fig. 14. The structure of spatial attention module (SAM).

$$M_C(F) = \sigma(MLP(\text{AvgPool}(F)) + MLP(\text{MaxPool}(F))) \quad (27)$$

The SAM module performs maximum and average pool operations on the input feature F' along the channel dimension, generates a two-layer feature matrix, and cascades them together. Then, a convolution kernel with a size of 7×7 is used to reduce the dimensions of the features, and the sigmoid function is applied to generate the spatial attention feature $M_S(F')$. The SAM calculation program can be expressed as:

$$M_S(F') = \sigma(f([\text{AvgPool}(F'); \text{MaxPool}(F')])) \quad (28)$$

where σ denotes the sigmoid function and f denotes the convolution operation with filter.

Table 2 lists the specific structural configuration of the proposed PV array fault diagnosis model based on CNN and CBAM. Convolution blocks with convolutional kernels of different sizes are stacked together along the depth direction to extract the features at different levels. This structure can automatically extract effective feature information directly from the original input feature matrix. In this study, two CBAM modules are respectively embedded into convolution modules, and the classification accuracy of the fault diagnosis model is improved by processing the focused feature information.

3.4. The application process of the proposed method in engineering

To detail the application process of the proposed method in engineering, Fig. 15 shows a schematic of a PV system using fault detection and diagnosis technology. Specifically, the PV array model consistent with the array structure and PV module parameters of an actual PV system is constructed using PSCAD/EMTDC, and the characteristic curves for each fault type are obtained as the dataset for fault diagnosis training purposes. Subsequently, the Isc-Voc GADF method is used to pre-process the data of the characteristic curves to train the proposed CNN-CBAM fault classification model. Furthermore, in actual PV plant operation, fault monitoring is carried out based on our previous work [64]. The I-V tracer is triggered to collect I-V/P-V curves after abnormal fault monitoring, and fault diagnosis is then performed using the proposed method.

4. Results and discussion

4.1. Experimental setup

The data used in this study are based on the configurations described in Section 3.1. Specifically, the faults in PV arrays with two blocking diode configurations under various operating conditions are analyzed, including 14 faults under contamination-prone operating conditions and 9 faults under ideal operating conditions. To ensure that the data fully reflect the actual operating conditions, we use the collected annual ambient records of the actual power station as the environmental control input for each fault type to obtain the corresponding characteristic curves. The data is divided into training and testing sets that accounted for 80% and 20% of the data, respectively. In addition, 90% of the training set comprise training data, and 10% comprise validation data. The data volume and proportion of each dataset applied to the different operating conditions are listed in Table 3.

The classification methods analyzed, including ANN- and CNN-based methods, are coded in the Keras framework of TensorFlow and are implemented using the structure of neural networks. The specific network structures of all the models used for comparison, including ANN, multilayer CNN, multi-scale CNN, and the proposed CNN-CBAM, as well as the range and combination of hyperparameters for the grid search of each model [65,66], are listed in the appendix.

4.2. Faults diagnosis evaluation indexes

The widely used precision, recall, F1-score, and accuracy are selected as indicators of the effectiveness of the classification algorithm [67,68], and can be calculated as:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (29)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (30)$$

$$\text{F1_score} = \frac{2 \cdot \text{Recall} \cdot \text{Precision}}{\text{Recall} + \text{Precision}} \quad (31)$$

Table 2
Detailed configuration of the CNN-CBAM.

Layer	Output shape	Detailed structure
Input Layer	(50, 50, 2)	
Convolution 1	(44, 44, 64)	$k = 3 \times 3$, filter = 8, $stride = 1 \times 1$, padding = 1 $k = 3 \times 3$, filter = 32, $stride = 1 \times 1$, padding = 1
CBAM 1	(44, 44, 64)	$k = 3 \times 3$, filter = 64, $stride = 1 \times 1$, padding = 1
Convolution 2	(42, 42, 32)	
CBAM 2	(42, 42, 32)	
Global Avgpool	32	
Output Layer	classes	

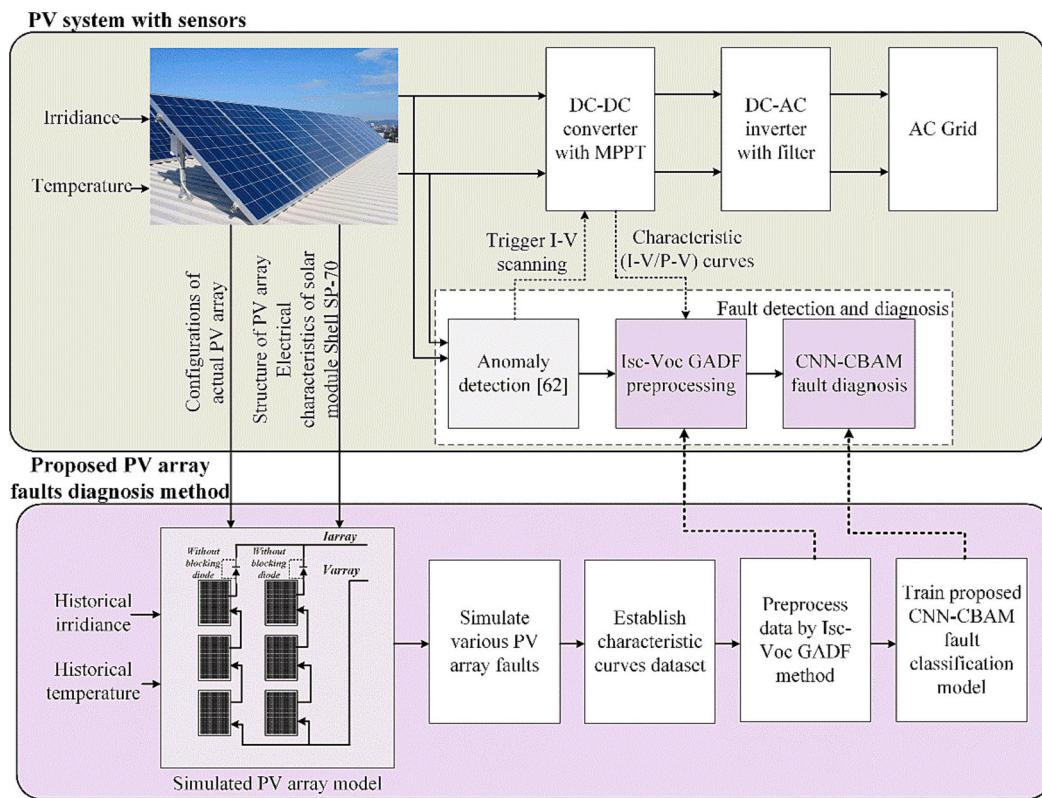


Fig. 15. Schematic diagram of a PV system with the proposed fault detection and diagnosis technology.

Table 3

The specific information of the different dataset.

	PV array with blocking diodes			PV array without blocking diodes		
	80%		20%	80%		20%
	Training data (90%)	Validation data (10%)	Testing data	Training data (90%)	Validation data (10%)	Testing data
Considering soiling impact	Number of classes	14	14	14	14	14
	Number of data	43,192	4800	11,998	43,192	4800
Without considering soiling impact	Number of classes	9	9	9	9	9
	Number of data	27,766	3086	7713	27,766	3086

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (32)$$

where the number of samples that belong to the positive category and are predicted to be in the positive category are referred to as true positives (TP); the number of samples that belong to the positive category and are predicted to be in the positive category are referred to as false negatives (FN); the number of samples that belong to the negative category and are predicted to be in the positive category are referred to as false positives (FP); and the number of samples that belong to the negative category and are predicted to be in the negative category are referred to as true negatives (TN).

4.3. Case 1. PV arrays under operating condition with non-uniform soiling impact

This section analyzes the effectiveness of different methods in distinguishing the 14 fault types of PV arrays under **Case 1** operating conditions with non-uniform soiling impact, including the accuracy and performance of different graphical feature transformation methods and

classification algorithms for fault diagnosis.

4.3.1. Analysis of graphical features transformation methods

First, we compare data pre-processing methods for fault diagnosis that do not rely on STC correction, as presented in [Section 2.2](#), including methods that combine I-V curves and environmental variables to form a two-dimensional data matrix (GTIV), and methods that utilize the complete distinguishable feature information of the characteristic curves. In addition, the diagnosis results of different graphical feature transformation methods such as GADF and recurrent plot (RP) using the complete characteristic curves are analyzed, and the three strategies of using the I-V graphical feature matrix, P-V graphical feature matrix, and IV-PV stacked graphical feature matrix obtained from the transformation as input are further compared.

[Table 4](#) presents the results of applying different data pre-processing methods to the characteristic curves for fault diagnosis, all of which use a multilayer-CNN classifier. When the graphical feature matrix transformed by single curve is used as the input, that is, the I-V or the P-V characteristic curve, the effectiveness of the information contained in the transformation matrix in distinguishing fault types is related to the

Table 4

The fault diagnosis results of applying different data preprocessing methods.

			Precision	Recall	F1-score	Accuracy
Circumstance 1: With blocking diodes	GADF	I-V	98.40%	98.31%	98.35%	98.46%
		P-V	95.27%	95.16%	95.21%	95.33%
		IV-PV	98.53%	98.48%	98.50%	98.58%
	Recurrent Plot	I-V	80.97%	80.03%	80.50%	81.77%
		P-V	86.87%	86.27%	86.57%	86.98%
	Direct IV	IV-PV	94.43%	94.23%	94.33%	95.12%
Circumstance 2: Without blocking diodes	GADF	GTIV	95.86%	95.79%	95.82%	95.98%
		I-V	98.38%	98.25%	98.31%	98.50%
		P-V	91.56%	91.32%	91.44%	92.06%
	Recurrent Plot	IV-PV	98.38%	98.28%	98.33%	98.48%
		I-V	67.04%	65.53%	66.28%	65.94%
	Direct IV	P-V	79.32%	78.25%	78.78%	78.69%
		IV-PV	90.08%	89.86%	89.97%	90.85%
		GTIV	95.61%	95.52%	95.56%	95.65%

feature transformation methods. Specifically, the graphical features of the GADF transformation are more effective for classification than the graphical features of the RP transformation. This is because the recurrent plot reconstructs the time series in binary (i.e. 0,1) format [69], whereas the GADF represents the time series in polar coordinates, where each element in the matrix is the cosine of the sum of angles. Therefore, the GADF transformation method contains more abundant and identifiable information about the characteristic curves than the information obtained by simply converting the RP into binary matrices. Additionally, for the fault classification of the PV array considering the impact of dust, the accuracies of the graphical feature matrix transformed by GADF are 98.58% and 98.48% in the configurations with and without blocking diodes, respectively, which are better than that of GTIV matrix methods. This is due to the fact that the transformed graph features by the proposed GADF effectively integrate G and T with the I-V/P-V curves, whereas the information of G and T in the latter GTIV is not effectively integrated into the features. Therefore, the completeness of valid information contained in the two features is different and the accuracy of GTIV diagnosis is limited. Moreover, the classification accuracy of the IV-PV stacked graphical feature matrix is typically higher than that of the single characteristic curve transformed matrix.

Fig. 16 shows the classification accuracy of the optimal feature strategy for the three pre-processing methods, where the stacked IV-PV matrix is used for the graphical feature transformation method. The accuracy of the RP transformation method for arrays with blocking

diodes is 95.12%, which is significantly higher than that for arrays without blocking diodes (90.85%). This may be related to the similarity between the two types of short-circuit and the health state without blocking the diode configuration when the RP transformation is applied, whereas the GADF transformation is applicable to different PV array configurations, with high accuracies of 98.58% and 98.48%, respectively.

The above comparative analysis confirms the effectiveness of the GADF graphical feature transformation method for fault diagnosis of PV arrays with different blocking diode configurations. Table 5 compares different GADF normalization approaches. The application of GADF with good diagnostic accuracy, first presented in [35] was based on the ANN method, while more types of faults and higher complexity of faults are involved in this study; therefore, the classification effects of multilayer-CNN and ANN are compared.

The results in Fig. 17 show that for both blocking diode configurations, the graphical feature transformation pre-processing method using the Isc-Voc normalized GADF is significantly more accurate in classifying faults than either the normal or global normalization strategies. In particular, when classifying with relatively simple methods such as ANN, the classification effect mainly depends on the effectiveness of the features in distinguishing faults. In the case of arrays with blocking diodes, for example, the classification accuracy of the Isc-Voc normalized GADF transformation is as high as 96.11%, whereas the highest classification accuracy of the other normalization strategies is only 91.42%. This also demonstrates that the Isc-Voc normalization strategy can obtain features that make the GADF transformation matrices of different fault types more distinguishable than other normalization strategies. Moreover, the classification accuracy of the CNN is significantly improved over that of the ANN when using the feature matrix of I-V or P-V obtained by normal normalization or global normalization strategies as input, with the diagnostic accuracies of the array without blocking diodes being improved by 10.56% and 14.83%, respectively. Furthermore, when the I-V/P-V matrix or stacked IV-PV matrix processed by the Isc-Voc normalized GADF is used as the input feature for the CNN model, they both have higher diagnostic accuracies, which is related to the stronger learning capability of the multilayer CNN.

The loss and accuracy of training and validation in Fig. 18 show that the IV-PV stacked feature matrix can quickly achieve a higher classification accuracy, indicating that the stacked matrix has the advantage of more significant differentiation than the others. Similarly, when the features processed by the Isc-Voc normalization strategy are used as the input for the classification model, the training accuracy of the model in Fig. 19 rapidly improves, and it takes less time to train to the highest accuracy than the features processed by other normalization strategies. In other words, the proposed Isc-Voc normalized GADF graphical feature matrix is more efficient and accurate when applied to fault diagnosis.

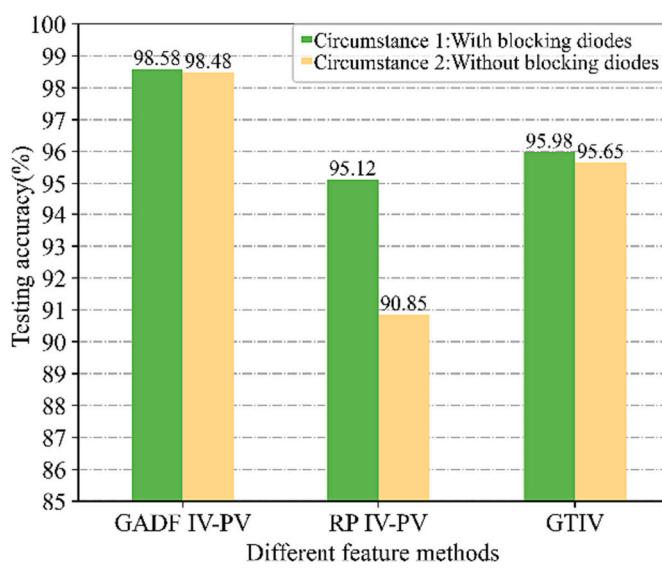
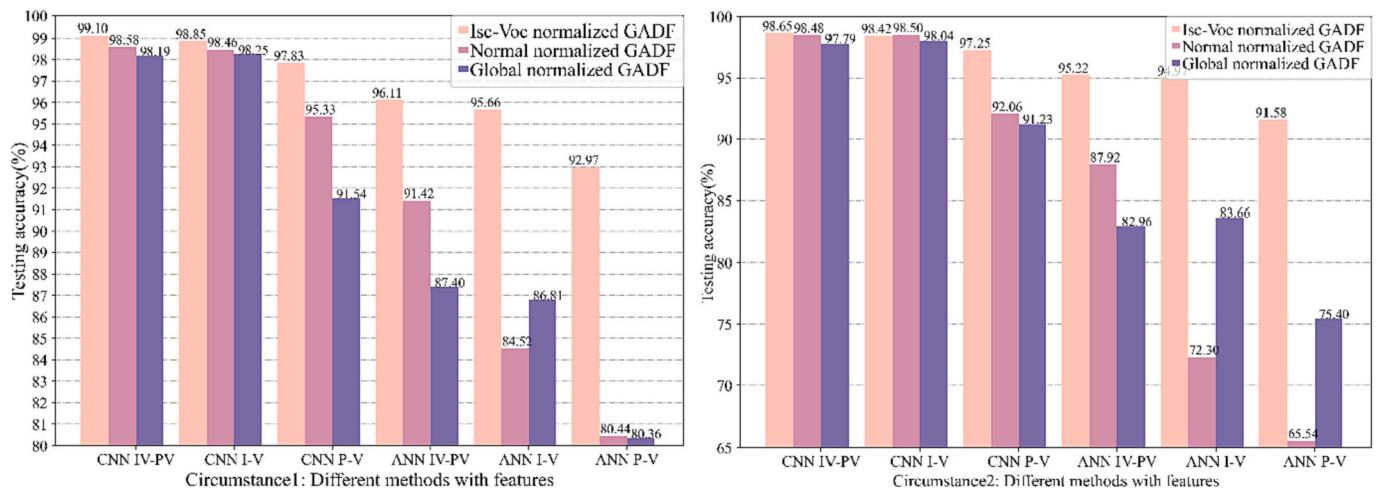
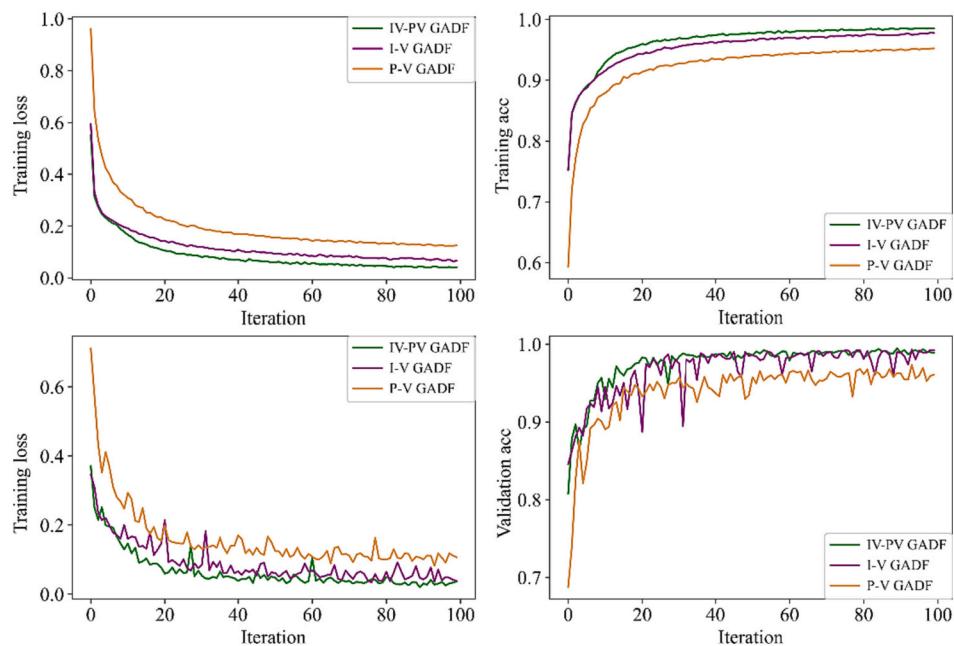


Fig. 16. Comparison of the optimal feature strategies for three pre-processing methods.

Table 5

The fault diagnosis results of applying different GADF normalization methods.

		GADF	Testing accuracy		
			Normal normalization	Global normalization	Isc-Voc normalization
Circumstance 1: With blocking diodes	ANN	IV	84.52%	86.81%	95.66%
		PV	80.44%	80.36%	92.97%
		IV-PV	91.42%	87.40%	96.11%
	MultilayerCNN	IV	98.46%	98.25%	98.85%
		PV	95.33%	91.54%	97.83%
		IV-PV	98.58%	98.19%	99.10%
Circumstance 2: Without blocking diodes	ANN	IV	72.30%	83.66%	94.97%
		PV	65.54%	75.40%	91.58%
		IV-PV	87.92%	82.96%	95.22%
	MultilayerCNN	IV	98.50%	98.04%	98.42%
		PV	92.06%	91.23%	97.25%
		IV-PV	98.48%	97.79%	98.65%

**Fig. 17.** Comparison of three GADF normalization methods with different input matrices and classifiers.**Fig. 18.** The loss and accuracy for three types of input matrices.

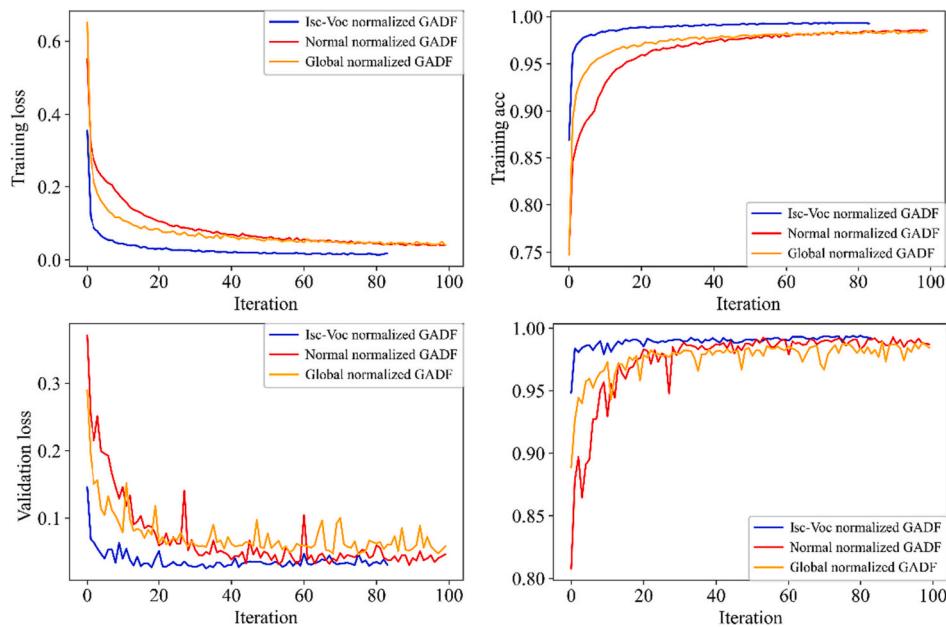


Fig. 19. The loss and accuracy for three GADF normalization methods.

4.3.2. Analysis of convolutional neural network classification models

Although the classification accuracy of the multilayer CNN described in the previous section is relatively considerable, the structure of this model is simple, and there is still room for improvement.

In this section, we further explore improvements in the classification accuracy of CNN-based models with different structures. Therefore, we design and compare the basic multilayer convolutional neural network, multi-scale convolutional neural network presented in [43], and the proposed CNN with the CBAM modules, whose structures are shown in Fig. 20.

The results in Table 6 show that the IV-PV stacked feature matrix obtained from the Isc-Voc normalized GADF transformation has the highest classification accuracy as the input to the CNN-based classification models for all three structures, regardless of the blocking diode configuration of the PV array. In addition, the multi-scale CNN performs better than the multilayer CNN when using the transformed features of all three GADF normalization strategies as inputs, mainly because the

multi-scale CNN extracts features at different scales. When the CBAM module is embedded in the multilayer CNN, the introduction of channel attention mechanism and spatial attention mechanism enable the network to extract features of interest in the local area while focusing on the channel features, which is more selective in focus than the multi-scale CNN with different scales of convolutional kernels for feature extraction. Overall, the CNN with the CBAM module has the highest classification accuracy, with the stacked IV-PV graphical feature matrix transformed by the proposed Isc-Voc normalized GADF as the input. The diagnostic accuracies applied to the PV arrays with the two diode configurations are 99.62% and 99.40%.

Furthermore, the pre-processing methods of the Isc-Voc normalized GADF and normal normalized GADF, which are both highly accurate when using CNN-CBAM as the classifier, are compared. Figs. 21 and 22 show the confusion matrices of the fault diagnosis for PV arrays with and without blocking diodes, respectively. For arrays with blocking diodes, the two normalization strategies are similar in terms of discrimination

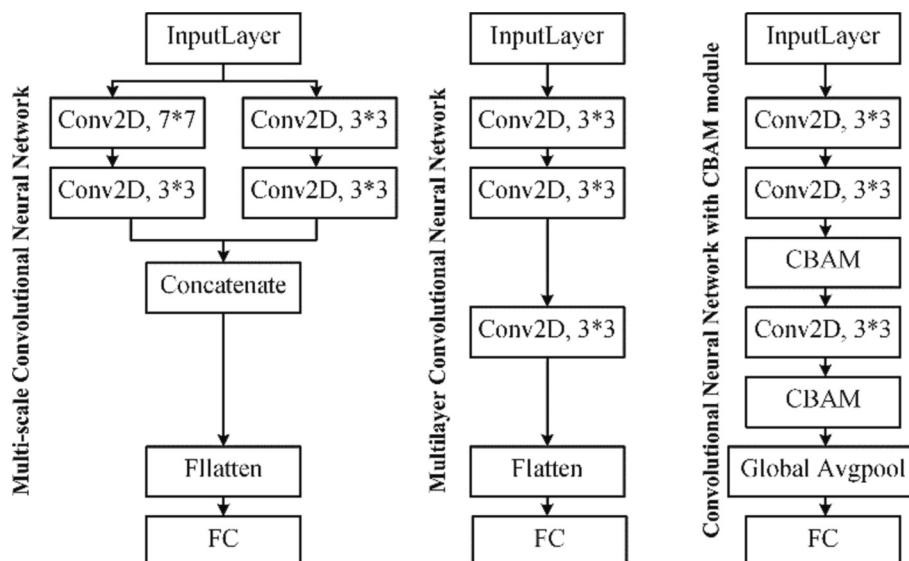


Fig. 20. The structures of compared three CNN-based classification models.

Table 6

The fault diagnosis results of applying different CNN-based classification models.

	CNN-based classifiers	Testing accuracy		
		Normal normalization	Global normalization	Isc-Voc normalization
Circumstance 1: With blocking diodes	Multilayer CNN	98.58%	98.19%	99.10%
	Multi-scale CNN	99.10%	98.65%	99.31%
	Proposed CNN-CBAM	99.42%	98.98%	99.62%
Circumstance 2: Without blocking diodes	Multilayer CNN	98.48%	97.79%	98.65%
	Multi-scale CNN	98.85%	98.54%	99.06%
	Proposed CNN-CBAM	99.15%	98.77%	99.40%

ability for most fault types. The specific difference is that the Isc-Voc normalized GADF performs slightly better than the normal normalized GADF in terms of OC, Adegradation, and Adelegation under soiling, owing to the similarity of these characteristic curves under certain environmental conditions and fault levels. The former reduces the proportion of OC misdiagnosed as Adegradation by 59% (13 samples) and does not produce 11 samples for which the latter diagnoses OC as Adegradation under soiling. In addition, for the arrays without blocking diodes, the characteristic curves of LL2 generated by the current backflow are highly similar to those of Adegradation and OC. The Isc-Voc normalized GADF has higher discrimination than the normal normalized GADF in this case, and their overall recall and F1-score both are 99.28% and 98.82%, respectively. Specifically, the Isc-Voc normalized GADF prevents 11 samples of LL2 from being misclassified as Adegradation, 15 samples of OC from being misidentified as Adegradation under soiling, and 11 samples of LL1 and OC under soiling that are not correctly distinguished. Moreover, the Isc-Voc normalized GADF is 27% more accurate than the normal normalized GADF in distinguishing between Adegradation under soiling

and LL2 under soiling and improved the diagnostic accuracy of complex faults affected by contamination by 41.94%.

Generally, the complexity of faults increases because of the impact of dust, and individual errors exist in the classification of all fault types. The proposed Isc-Voc normalized GADF method achieves the best accuracy for distinguishing complex faults. Its performance advantage is reflected in its ability to rapidly converge during the training process. In addition, the application of the transformed graphical features facilitates the rapid learning of parameters, providing the best classification performance when applied to the proposed CNN-CBAM, as shown in Fig. 23.

4.3.3. Statistical tests

Because several algorithms were compared, it was necessary to analyze whether there were significant differences between them. Therefore, it is necessary to perform statistical tests [70,71]. We performed ten-fold cross-validation on the training set, as shown in Table 7. The nonparametric Friedman test was applied to demonstrate the significant differences between the results obtained in the experiments [72], followed by a nonparametric Wilcoxon rank-sum test [73]. The Wilcoxon rank sum test at 0.05 confidence level was used to show significant differences between the different methods.

Table 8 presents the results of pairwise algorithm comparisons using the Wilcoxon test. The first two rows contain the values of the sum of the positive (R+) and negative (R-) ranks for each established comparison. The next two rows show the statistical values T and p for the significance levels. The last row indicates which algorithm wins for each comparison. The numbers of times each algorithm won are listed in Table 9. The results of the test clearly show that the Isc-Voc normalized CNN-CBAM receives the best ranking, followed by the Isc-Voc normalized multiscale CNN and normal normalized CNN-CBAM, and the statistical results further proving that the proposed method significantly outperforms the other algorithms.

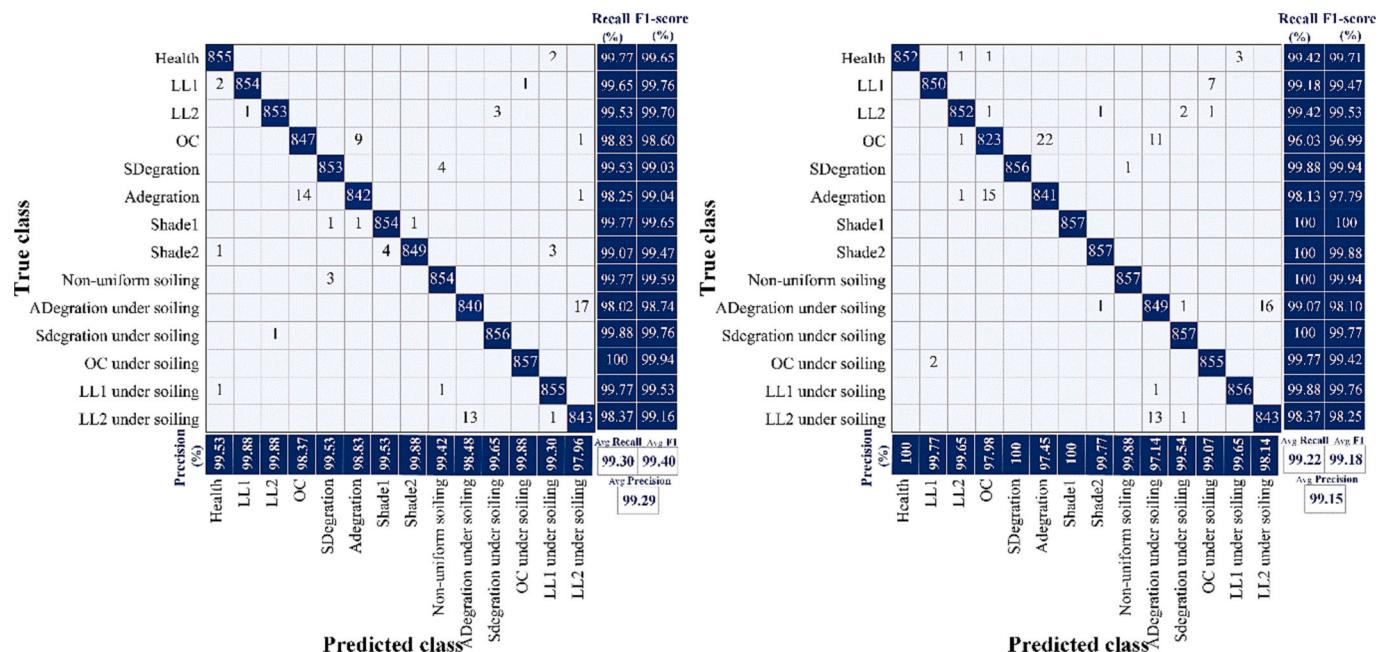


Fig. 21. The confusion matrices of PV arrays with blocking diodes under soiling condition.

The figure consists of two confusion matrices, (a) and (b), showing the performance of fault diagnosis methods under soiling conditions. Both matrices have 'True class' on the y-axis and 'Predicted class' on the x-axis. The columns are labeled 'Precision (%)' and 'Recall F1-score (%)'. The rows are labeled 'Avg Precision' and 'Avg Recall Avg F1'.

		Precision (%)		Recall F1-score (%)	
True class	Predicted class				
Health	Health	855	1	1	99.77 99.88
	LL1	853			99.53 99.53
	LL2	1	853		99.53 99.76
	OC		842	1 12	98.24 98.24
	SDegradation		853	1	99.53 99.35
	Adegredation		15	840	98.02 98.19
	Shade1			856	99.88 99.82
	Shade2			2 854	99.65 99.59
	Non-uniform soiling		3		99.65 99.65
	ADegredation under soiling				98.60 98.42
Sdegredation under soiling				100 99.82	
OC under soiling		3	2	99.30 99.41	
LL1 under soiling				100 99.94	
LL2 under soiling			15	98.25 98.30	
Avg Precision		99.28	Avg Recall Avg F1		99.28 99.28

		Precision (%)		Recall F1-score (%)	
True class	Predicted class				
Health	Health	853	3	1	99.53 99.76
	LL1	845			98.60 98.77
	LL2	836	3	11	97.55 98.29
	OC	833	1	8	97.20 97.83
	SDegradation			857	100 99.94
	Adegredation		8	9	97.90 97.85
	Shade1			857	100 100
	Shade2			857	100 99.94
	Non-uniform soiling				100 99.48
	ADegredation under soiling				832 2 23
Sdegredation under soiling				9 1 847	
OC under soiling		6		99.30 98.95	
LL1 under soiling				1 851 5	
LL2 under soiling			14 1	98.25 97.51	
Avg Precision		98.82	Avg Recall Avg F1		98.82 98.82
Avg Precision		98.83	Avg Precision		98.83

(a) Fault diagnosis using Isc-Voc normalized GADF

(b) Fault diagnosis using normal normalized GADF

Fig. 22. The confusion matrices of PV arrays without blocking diodes under soiling condition.

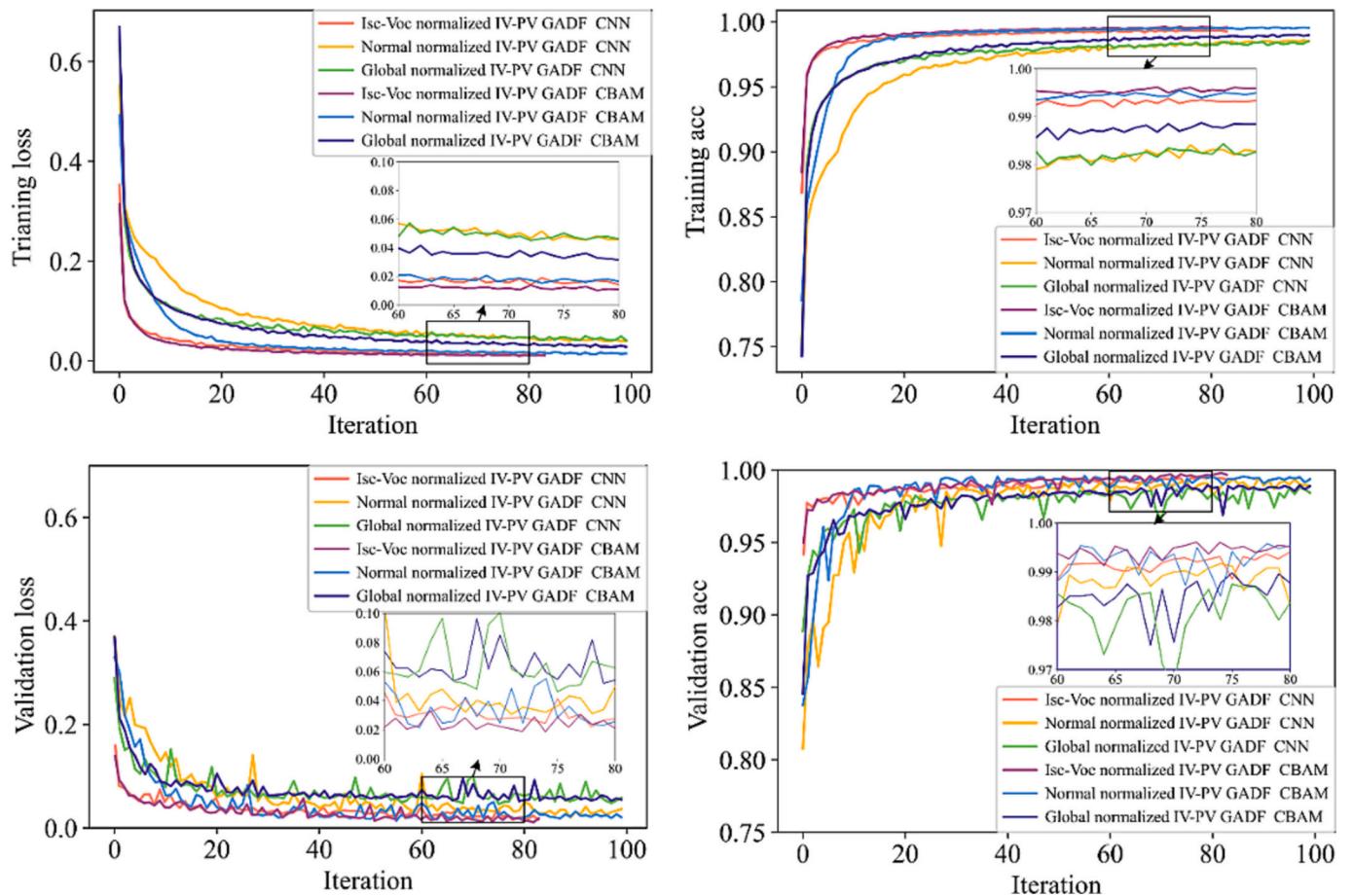


Fig. 23. The loss and accuracy for typical CNN-based classifiers with three GADF normalization methods.

Table 7

Cross-Validation test results at condition considering soiling effect.

		Isc-Voc normalization			CNN-CBAM		
		CNN-CBAM	Multi-scale CNN	Multilayer CNN	Isc-Voc norm	Normal norm	Global norm
Circumstance 1: With blocking diodes	Mean \pm S.	0.99334 \pm 0.00147	0.99059 \pm 0.00107	0.98800 \pm 0.00280	0.99334 \pm 0.00147	0.99195 \pm 0.00178	0.98309 \pm 0.00519
Circumstance 2: Without blocking diodes	Mean \pm S.	0.99101 \pm 0.00155	0.98706 \pm 0.00112	0.98352 \pm 0.00302	0.99101 \pm 0.00155	0.98897 \pm 0.00181	0.98166 \pm 0.00522

Table 8

Statistical results based on the Wilcoxon rank sum test for condition considering dust impact.

		Isc-Voc normalization			CNN-CBAM		
		CNN-CBAM Vs Multi-scale CNN	CNN-CBAM Vs Multi-scale CNN	Multi-scale CNN Vs Multilayer CNN	Isc-Voc norm Vs Normal norm	Isc-Voc norm Vs Global norm	Normal norm Vs Global norm
Circumstance 1: With blocking diodes	$\sum R^+$	55	55	52.5	48	54	55
	$\sum R^-$	0	0	2.5	7	1	0
	T	0	0	2	5	1	0
	p-value	0.00506	0.00506	0.01516	0.03815	0.00691	0.00506
Circumstance 2: Without blocking diodes	Winner	CNN-CBAM	CNN-CBAM	Multi-scale CNN	Isc-Voc norm	Isc-Voc norm	Normal norm
	$\sum R^+$	55	55	48.5	48	55	55
	$\sum R^-$	0	0	6.5	7	0	0
	T	0	0	5	5	0	0
	p-value	0.00506	0.00506	0.03815	0.03815	0.00506	0.00506
	Winner	CNN-CBAM	CNN-CBAM	Multi-scale CNN	Isc-Voc norm	Isc-Voc norm	Normal norm

Table 9

Rank of the comparison between the experiments.

	Methods	No. Wins	Ranking
Circumstance 1: With blocking diodes	Isc-Voc normalization	CNN-CBAM	4
		Multi-scale CNN	2
		Multilayer CNN	1
		Isc-Voc norm	4
	Proposed CNN-CBAM	Normal norm	2
		Global norm	0
		CNN-CBAM	4
		Multi-scale CNN	2
	Isc-Voc normalization	Multilayer CNN	1
		Isc-Voc norm	4
		Normal norm	2
		Global norm	0
Circumstance 2: Without blocking diodes	Proposed CNN-CBAM	Normal norm	2
		Global norm	0
		Normal norm	2
		Global norm	5

4.4. Case 2. PV arrays under operating condition without non-uniform soiling impact

4.4.1. Analysis of classification models

To further illustrate the universality of the proposed method, the fault diagnosis performance under normal operating conditions without

Table 10

The fault diagnosis results of applying ANN with different GADF methods.

	GADF	Testing accuracy		
		Normal normalization	Global normalization	Isc-Voc normalization
Circumstance 1:	IV	87.59%	93.20%	97.57%
With blocking diodes	PV	93.26%	91.02%	97.57%
	IV-PV	98.09%	93.68%	98.26%
Circumstance 2:	IV	75.28%	94.44%	97.81%
Without blocking diodes	PV	82.74%	91.34%	97.89%
	IV-PV	94.44%	94.83%	98.22%

dust impact is analyzed, which contains 9 fault states. Similarly, we analyze the influence of the transformed graphical features processed by various GADF normalization methods as inputs on the accuracy of the classification algorithms. Specifically, as shown in [Table 10](#), for simple classifiers such as ANN, using the matrix of I-V or P-V curves converted by Isc-Voc normalized GADF as input results in a significant classification accuracy improvement of 97.57%, compared with 93.26% for normal normalized GADF and 93.20% for global normalized GADF. The accuracy of the proposed normalized GADF reaches 98.26% when the IV-PV stacked transformation matrices are used as the input features. Moreover, as shown in [Table 11](#), the features processed by the optimal GADF strategy, that is, stacked IV-PV transformed by Isc-Voc normalization, exhibit an accuracy advantage of 99.84% when applied to CNN-CBAM. This is substantially better than the 98.26% over 1.58% for the ANN, and higher than the 99.58% over 0.26% for the multilayer CNN, which is smaller than the 3.51% advantage for the ANN and the 0.52% advantage for the multilayer CNN when the effect of dust is considered. This is because the types of faults to be distinguished during operation are not diverse, and their complexity is general. In other words, for faults not affected by dust, the graphical feature matrices transformed by Isc-Voc normalized GADF applied to a relatively simple structured CNN can

Table 11

The fault diagnosis results of applying classifiers with different GADF normalization methods.

	Classifiers	Testing accuracy		
		Normal normalization	Global normalization	Isc-Voc normalization
Circumstance 1: With blocking diodes	ANN	98.09%	93.68%	98.26%
	Multilayer CNN	99.42%	98.54%	99.58%
	Proposed CNN-CBAM	99.18%	98.35%	99.84%
	ANN	94.44%	94.83%	98.22%
Circumstance 2: Without blocking diodes	Multilayer CNN	99.22%	98.61%	99.42%
	Proposed CNN-CBAM	99.25%	98.38%	99.81%

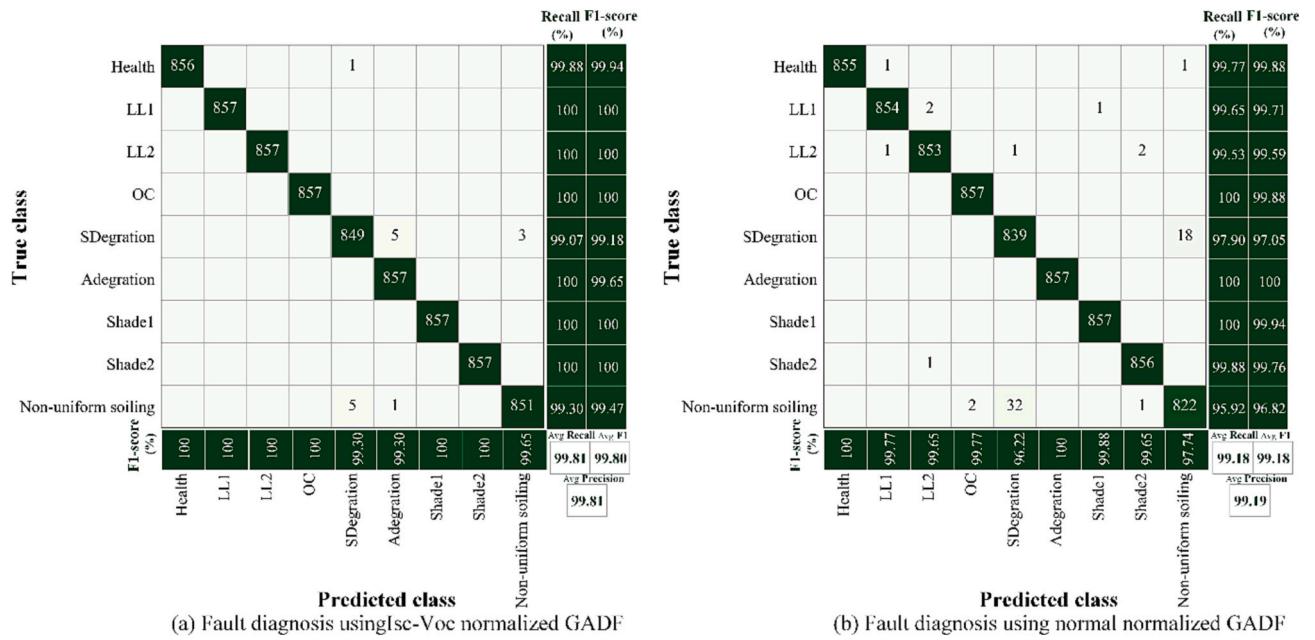


Fig. 24. The confusion matrix of PV arrays with blocking diodes.

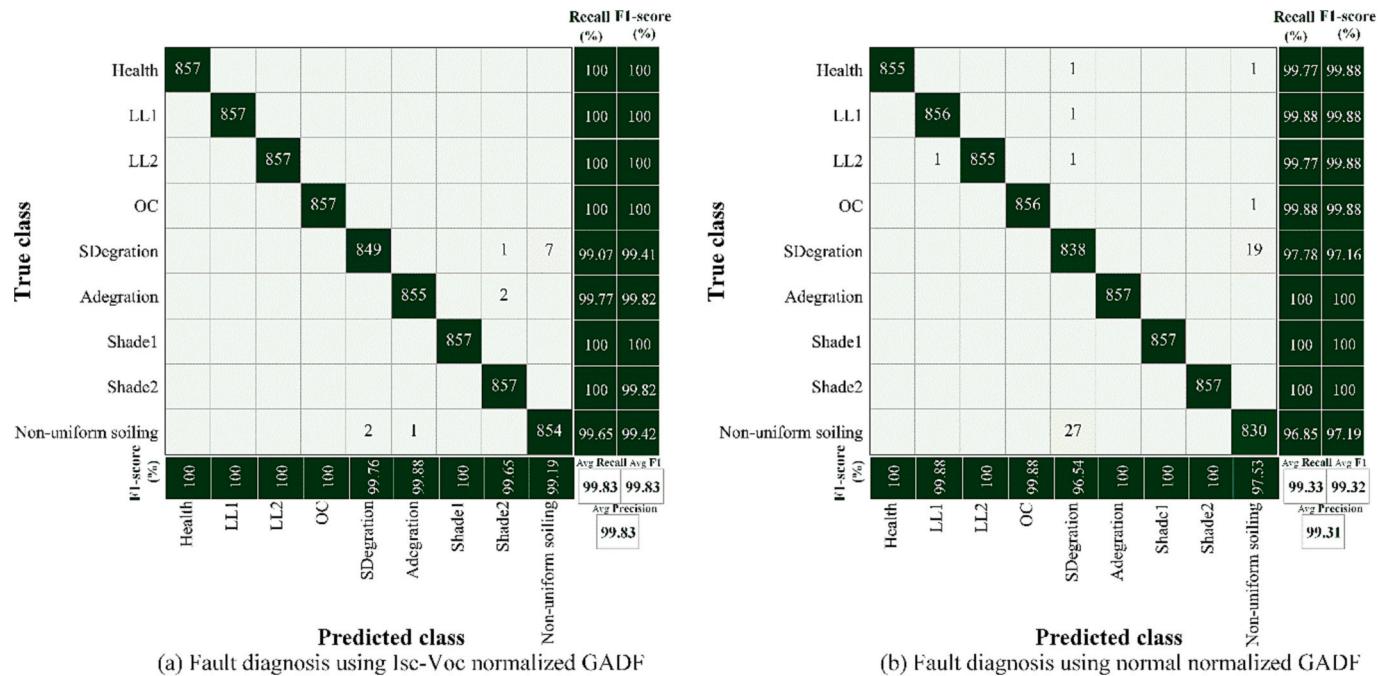


Fig. 25. The confusion matrix of PV arrays without blocking diodes.

obtain satisfactory results, but the proposed CNN-CBAM method still has the advantage of the highest accuracy.

Figs. 24 and 25 show the fault classification results of the Isc-Voc normalized GADF and normal normalized GADF, respectively, for the PV array with two blocking diode configurations. The confusion matrix of the former does not contain many sporadic misclassifications of the latter and is expressed with the overall recall and F1-score of 99.81% and 99.18%, respectively.

In the fault diagnosis of the PV array with blocking diode configuration, the classification error between SDegradation and Non-uniform soiling is 84% less in the former than in the latter, with 8 and 50 samples, respectively, and the proportion is 80.43% in the array without

blocking diode. This may be due to the low fault levels of these two fault types with variable fault parameters, that is, low fault differentiation between some SDegradations with low fault levels and low degrees of dust shielding. In fact, the correct classification accuracy for both SDegradation and Non-uniform soiling, which may be misclassified, still reaches over 99% for either diode configuration of the array. Overall, the proposed CNN-CBAM model based on Isc-Voc normalized GADF transformation can achieve high accuracy in fault classification and is also applicable to ideal operating conditions without dust impact.

The training and validation processes for the two blocking-diode configurations are shown in Figs. 26 and 27, respectively. When the inputs are graphical features processed by Isc-Voc normalized GADF, the

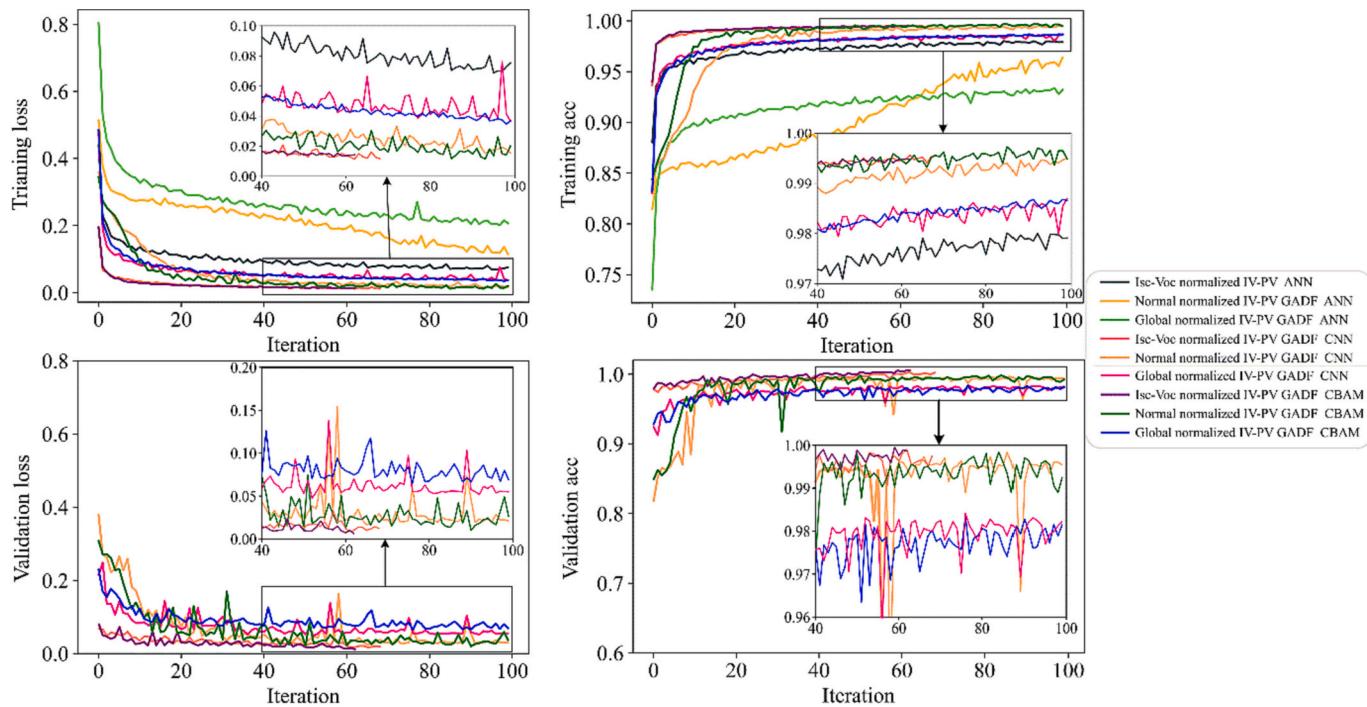


Fig. 26. The loss and accuracy of PV arrays with blocking diodes.

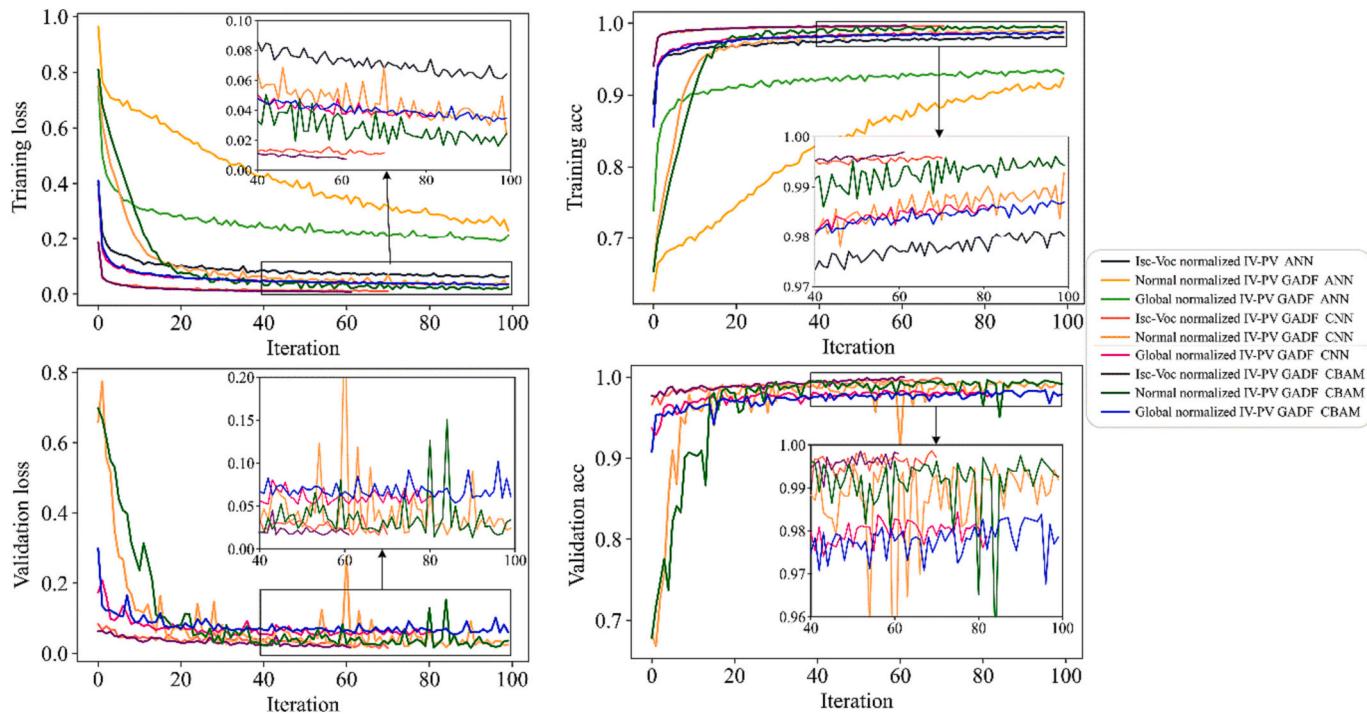


Fig. 27. The loss and accuracy of PV arrays without blocking diodes.

loss of the model decreases the fastest and the accuracy increases rapidly compared to other normalization strategies, whether ANN, CNN, or CNN-CBAM are used as classifiers. This is consistent with the original intention that the proposed graphical feature transformation method can achieve high discrimination between similar faults in different environments.

In particular, for arrays without blocking diode configurations, the model that classifies similar fault types using graphical features transformed by normal normalization strategy requires a relatively long

learning stage, and the model is not as stable as models using other strategies. This is because the classification accuracy relies heavily on the feature learning capability of the classifiers when the discriminative ability of features is constrained. Moreover, in terms of the classification ability of models with different levels of complexity, simple models such as ANN require longer training iterations than CNN-based models, and the proposed Isc-Voc normalized GADF can significantly improve the diagnosis performance when applied to the category of simple classifiers. For faults not affected by dust, CNN-CBAM still has a slightly better

Table 12

Cross-Validation test results at condition without considering soiling effect.

	Mean \pm S.	Isc-Voc normalization			CNN-CBAM		
		CNN-CBAM	Multi-layer CNN	ANN	Isc-Voc norm	Normal norm	Global norm
Circumstance 1: With blocking diodes	Mean \pm S. D.	0.99717 \pm 0.00141	0.99342 \pm 0.00107	0.98800 \pm 0.00331	0.99717 \pm 0.00141	0.99610 \pm 0.00136	0.98383 \pm 0.00424
Circumstance 2: Without blocking diodes	Mean \pm S. D.	0.99690 \pm 0.00149	0.99181 \pm 0.00115	0.98724 \pm 0.00347	0.99690 \pm 0.00149	0.99536 \pm 0.00144	0.98440 \pm 0.00433

Table 13

Wilcoxon rank sum test for condition without considering dust impact.

		Isc-Voc normalization			CNN-CBAM		
		CNN-CBAM Vs ANN	CNN-CBAM Vs Multilayer CNN	Multilayer CNN Vs ANN	Isc-Voc norm Vs Normal norm	Isc-Voc norm Vs Global norm	Normal norm Vs Global norm
Circumstance 1: With blocking diodes	$\sum R^+$	55	54	55	43	55	55
	$\sum R^-$	0	1	0	9	0	0
	T	0	0	0	0	0	0
	p-value	0.00503	0.00503	0.00503	0.01162	0.00503	0.00503
Circumstance 2: Without blocking diodes	Winner	CNN-CBAM	Multilayer CNN	Multilayer CNN	Isc-Voc norm	Isc-Voc norm	Normal norm
	$\sum R^+$	55	55	55	42	55	55
	$\sum R^-$	0	0	0	12	0	0
	T	0	0	0	0	0	0
	p-value	0.00503	0.00503	0.0053	0.01162	0.0053	0.0053
	Winner	CNN-CBAM	Multilayer CNN	Multilayer CNN	Isc-Voc norm	Isc-Voc norm	Normal norm

Table 14

Rank of the comparison between the experiments.

	Methods	No. Wins	Ranking
Circumstance 1: With blocking diodes	CNN-CBAM	4	1
	Multilayer	2	2
	CNN	0	4
	Isc-Voc norm	4	1
	Proposed CNN-CBAM	Normal norm	3
	Global norm	0	4
	CNN-CBAM	4	1
	Isc-Voc normalization	Multilayer	2
Circumstance 2: Without blocking diodes	CNN	0	4
	ANN	0	4
	Isc-Voc norm	4	1
	Proposed CNN-CBAM	Normal norm	3
	Global norm	0	4
	CNN	2	2
	ANN	0	4
	Isc-Voc norm	4	1

accuracy and training performance than the multilayer CNN.

4.4.2. Statistical tests

We applied a ten-fold cross-validation on the training set in [Table 12](#) and applied the Wilcoxon rank-sum test to the comparison results in [Table 13](#) to demonstrate the significant differences in the proposed method.

The numbers of wins for each algorithm are listed in [Table 14](#). The test results clearly show that the Isc-Voc normalized CNN-CBAM obtains the best ranking, followed by the Isc-Voc normalized multilayer CNN, and the statistical results further prove that the proposed method does outperform the other algorithms.

4.5. Discussion

Compared with the most advanced research in the literature, our approach enables the accurate identification of variable fault parameter defects for arrays with multiple scenarios and configurations. Among the

categories using GADF-transformed graphical features as classification features, the IV-PV stacked matrix by the proposed Isc-Voc normalized GADF pre-processing method outperforms the transformed features of single characteristic curve in [35], both in terms of classification accuracy and model training performance. This implies that the stacked GADF graphical features can highlight the distinguishability of different fault classes in pre-processing methods without employing correction of field characteristic curves, and that the GADF based on Isc-Voc normalization has a more reliable discriminatory capability. The main reasons for this are that stacked graphical features contain richer fault information than features of single curve and that the Isc-Voc normalized GADF overcomes the challenge of inconsistent feature characterization of the same faults in different environments. In the comparative analysis of different classifiers, the proposed CNN-CBAM has a certain accuracy advantage over other CNN-based methods and is superior to simple models such as ANN, particularly under dust-influenced operating conditions containing multiple mixed and high-complexity variable fault parameters. This is mainly because the characteristic curves jointly affected by the variable fault parameters and changing environmental factors are more complex, and the learning ability of simple classifiers is limited. Notably, the input features have a greater impact on classification performance, both in terms of accuracy and robustness, than the type of classifier used. This notion has been proven in terms of the significantly better diagnostic performance of Isc-Voc normalized GADF features compared with other GADF features using a simple ANN model as the classifier. Furthermore, the benefits of the proposed Isc-Voc normalized GADF are still notable and effective when the impact of dust is not considered.

These results prove that the features are extremely important; that is, the original data converted into effective features can improve the discriminatory quality of the input features. It can simplify the tuning process of the classifier and improve the diagnostic performance. Moreover, the CNN structure with the introduction of the CBAM attention module based on high-quality effective features offers significant advantages for the diagnosis of complex fault types.

5. Conclusion

In this study, we propose a new fault diagnosis method for identifying and evaluating faults in PV arrays, which can cope with PV arrays

with different blocking diode configurations under both operating conditions, with and without considering the impact of dust. The salient aspects of our approach are summarised as follows. To facilitate the extraction of full information in the characteristic curves by the classifier, we transform both the I-V and P-V curves into a graphical feature matrix of the GADF and stack them together. Then, for reliable and easy-to-implement practical applications that require the features of the same faults in different environments to be effectively uniform, we adopt calibration-independent field characteristic curves and normalized them by GADF using the ideal I_{sc} and V_{oc} under their ambient conditions. Furthermore, to overcome the challenge of classifiers for multiple variable fault parameters in complex operating conditions, we employ CNN embedded with CBAM modules to identify and classify the graphical features of complicated faults.

In various scenarios, the performance of the proposed technology was verified through statistical analysis of cross-validation experiments in terms of accuracy and stability. Specifically, the I_{sc} - V_{oc} normalized GADF-transformed graphical features have higher fault discrimination than the features transformed by the other strategies, significantly simplifying the processing of the classifier and obtaining excellent diagnostic accuracy for both conditions, with or without dust effects. And it is verified that CNN-CBAM is effective for fault identification under normal operating conditions with relatively simple faults or dust operating conditions with complex faults. In general, the proposed method based on GADF-transformed graphical features of characteristic curves and convolutional neural network with CBAM modules is applicable to PV arrays with or without blocking diodes and is also effective for fault diagnosis under different operating conditions, including concurrent faults affected by dust. They can improve the efficiency of operation and maintenance, reduce fault time, and provide economic benefits to power stations. The proposed I_{sc} - V_{oc} normalized GADF transformation provides a new scheme for using full characteristic curve information by discarding STC correction and relieving the experimental dependence on correction factors, which greatly expands the universal application of fault diagnosis. Based on this research, there are

still challenges that need to be addressed in future work, such as further differentiation of different dynamic fault levels and the location of strings that occur during diagnosis, which will play an important role in improving the O&M efficiency and return on investment. More importantly, the proposed fault diagnosis technology can be applied in conjunction with the abnormal monitoring of PV array operations in online power plant engineering.

CRediT authorship contribution statement

Jiaqi Qu: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Zheng Qian:** Supervision, Funding acquisition, Conceptualization. **Qiang Sun:** Writing – review & editing, Writing – original draft, Methodology. **Lu Wei:** Conceptualization. **Hamidreza Zareipour:** Writing – review & editing, Validation, Conceptualization.

Declaration of Competing Interest

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

Data availability

The data that has been used is confidential.

Acknowledgement

This work is supported by the National Natural Science Foundation of China (No. 61573046), Program for Changjiang Scholars and Innovative Research Team in University (No. IRT1203), and the Academic Excellence Foundation of BUAA for PhD Students.

Appendix A. Appendix

To demonstrate that the proposed method is effective for data containing noise, we add 50 and 35 dB of noise to the characteristic curves, as shown in Fig. A1. Table A1 lists the diagnostic results of the proposed CNN-CBAM method on the graphical features of noisy characteristic curves using the I_{sc} - V_{oc} normalized GADF-transformed method (Example of a PV array structure configuration containing blocking diodes). The results indicate that for the noise data with 35 dB added, the diagnostic accuracies of the method proposed in this study are 92.42% and 95.50% for the two working conditions, regardless of whether they are affected by soiling; for the data with 50 dB noise added, the diagnostic accuracies are 98.45% and 99.17%, respectively. It can be confirmed from Figs. A2–A3 that the method proposed in this study is effective for the diagnosis of faults containing noisy data.

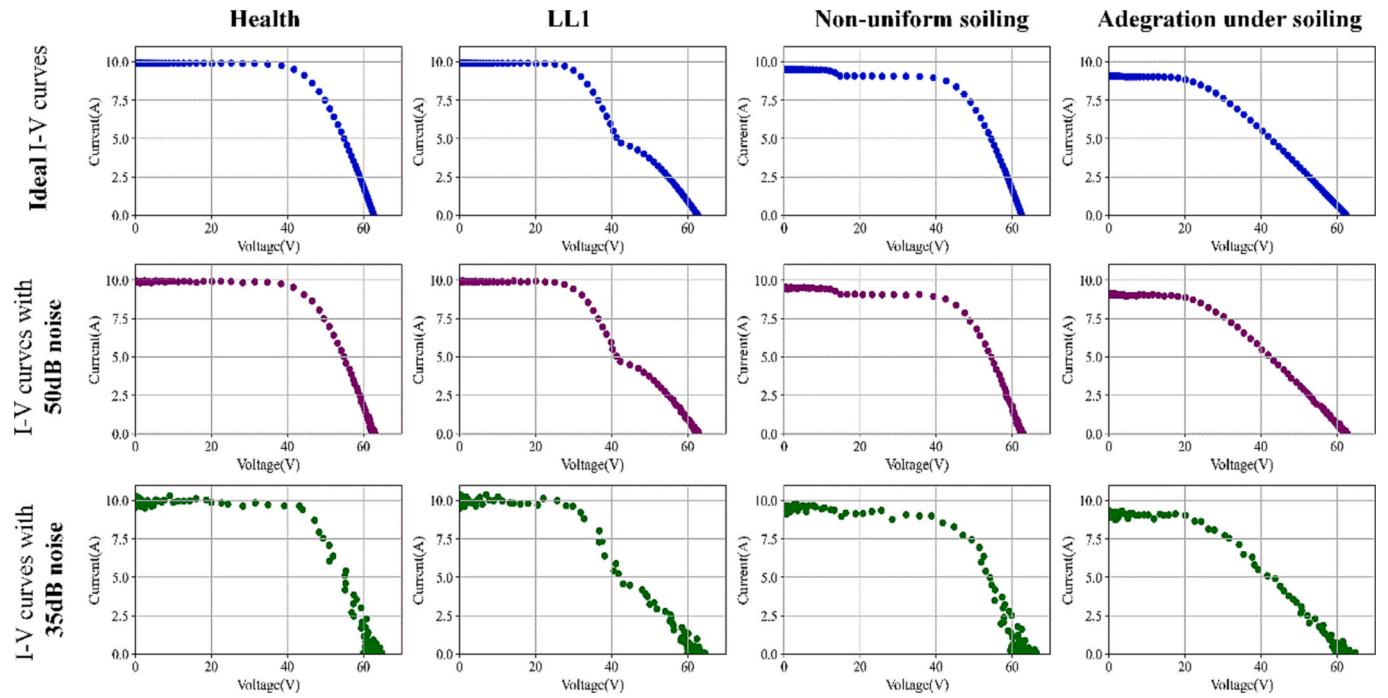


Fig. A1. I-V curves with noise.

Table A1

Diagnostic results of the proposed CNN-CBAM method with noisy characteristic curves.

	50 dB	35 dB
Condition considering soiling effect	98.45%	92.42%
Condition without considering soiling effect	99.17%	95.50%

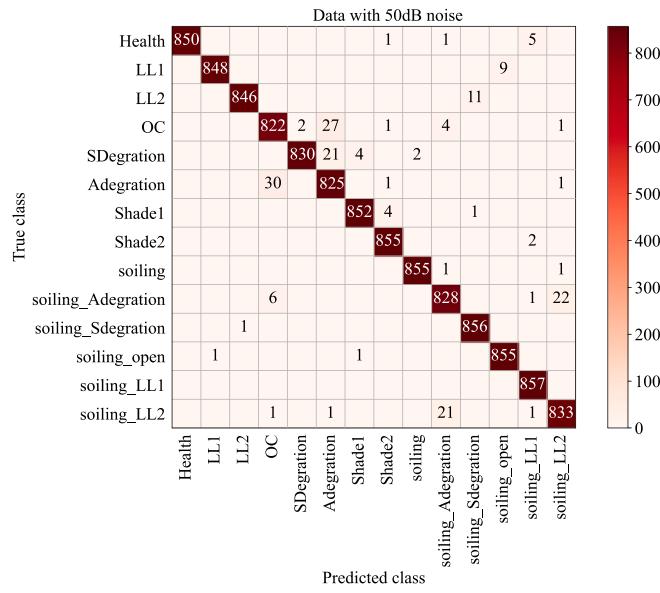


Fig. A2. The confusion matrices of PV arrays with non-uniform soiling impact.

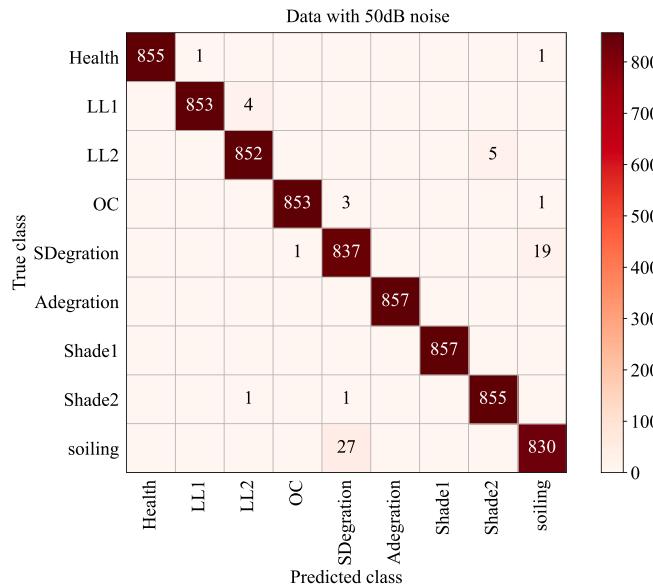


Fig. A3. The confusion matrices of PV arrays without non-uniform soiling impact.

Tables A2–A5 are the grid search ranges for the hyperparameters of the models compared in paper, including ANN, multilayer CNN, multi-scale CNN, and the proposed CNN-CBAM. And Tables A6–A9 are the specific network structures of the above models. The classification layer is determined by the number of fault types for which dust effects are considered or not, i.e., 14 types for which dust is considered and 9 types for which dust is not considered. The specific experimental configuration was performed on an NVIDIA GeForce GTX 2080 Ti, and the network was built with Tensorflow-gpu 1.5.0, Keras 2.1.6. The Adam optimizer was used, and the epoch = 64, batch size = 300, and early stopping strategy with patience = 20 were selected to ensure that the parameters of model were adequately trained.

Table A2

Grid search range for ANN hyperparameters.

	Dense 1	Dense 2	Dense 3	classification
ANN	Unit1 = [20, 25, 30, 40, 50]		/	[9/14]
	Unit1 = [20, 25, 30, 40, 50]	Unit2 = [10, 15, 20, 30, 40, 50]	/	[9/14]
	Unit1 = [20, 25, 30, 40, 50]	Unit2 = [10, 15, 20, 30, 40, 50]	Unit3 = [10, 15, 20, 30, 40, 50]	[9/14]

Table A3

Grid search range for Multilayer CNN hyperparameters.

	Conv2D 1	Conv2D 2	Dense 1	classification
Multilayer CNN	Filter1= [5, 10, 16, 32, 64]	Filter2= [5, 10, 16, 32, 64]	Unit= [10, 16, 25, 32, 40, 50]	[9/14]
	Kernel_size1= [(3,3), (5,5), (7,7)]	Kernel_size2= [(3,3), (5,5), (7,7)]		
	Stride1= [(1,1), (2,2)]	Stride2= [(1,1), (2,2)]		

Table A4

Grid search range for Multi-scale CNN hyperparameters.

	Conv2D 1	Conv2D 2	Conv2D 3	Dense 1	classification
Multi-scale CNN	Filter1= [5, 10, 16, 32, 64]	Filter2= [5, 10, 16, 32, 64]	Filter3= [5, 10, 16, 32, 64]	Unit= [10, 16, 25, 32, 40, 50]	[9/14]
	Kernel_size1= [(3,3), (5,5), (7,7)]	Kernel_size2= [(3,3), (5,5), (7,7)]	Kernel_size3= [(3,3), (5,5), (7,7)]		
	Stride1= [(1,1), (2,2)]	Stride2= [(1,1), (2,2)]	Stride3= [(1,1), (2,2)]		

Table A5

Grid search range for Proposed CNN-CBAM hyperparameters.

	Conv2D 1	Conv2D 2	CBAM 1	Conv2D 3	CBAM 2	classification
Proposed CNN-CBAM	Filter1= [16, 32, 64, 128]	Filter2= [16, 32, 64, 128]	spatial_attention kernel_size1= [(3,3), (7,7)]	Filter3= [16, 32, 64, 128]	spatial_attention kernel_size2= [(3,3), (7,7)]	[9/14]
	Kernel_size1= [(3,3), (5,5), (7,7)]	Kernel_size2= [(3,3), (5,5), (7,7)]	channel_attention reduction_ratio1= [0.3,0.5]	Kernel_size3= [(3,3), (5,5), (7,7)]	channel_attention reduction_ratio2= [0.3,0.5]	
	Stride1= [(1,1), (2,2)]	Stride2= [(1,1), (2,2)]		Stride3= [(1,1), (2,2)]		

Table A6

Specific structure of ANN.

Layer (type)	Output Shape	Number of parameters	Connected to	Total parameters
input_a (InputLayer)	(None, 50, 50, 2)	0		151,844
flatten_1 (Flatten)	(None, 5000)	0	input_a[0][0]	
dense_1 (Dense)	(None, 30)	150,030	flatten_1[0][0]	
dense_2 (Dense)	(None, 40)	1240	dense_1[0][0]	
dense_3 (Dense)	(None, 14)	574	dense_2[0][0]	

Table A7

Specific structure of Multilayer CNN.

Layer (type)	Parameter configuration	Output Shape	Number of parameters	Connected to	Total parameters
input_a (InputLayer)		(None, 50, 50, 2)	0		627,209
conv2d_1 (Conv2D)	Filter1= [10] Kernel_size1 = [(3,3),] Stride1 = [(1,1)]	(None, 50, 50, 10)	190	input_a[0][0]	
conv2d_2 (Conv2D)	Filter2= [5] Kernel_size2 = [(5,5),] Stride2 = [(1,1)]	(None, 50, 50, 5)	1255	conv2d_1[0][0]	
dropout_1 (Dropout)		(None, 50, 50, 5)	0	conv2d_2[0][0]	
flatten_1 (Flatten)		(None, 12,500)	0	dropout_1[0][0]	
dense_1 (Dense)	Unit= [50]	(None, 50)	625,050	flatten_1[0][0]	
dense_2 (Dense)		(None, 14)	714	dense_1[0][0]	

Table A8

Specific structure of multi-scale CNN.

Layer (type)	Parameter configuration	Output Shape	Number of parameters	Connected to	Total parameters
input_a (InputLayer)		(None, 50, 50, 2)	0		818,424
conv2d_1 (Conv2D)	Filter1= [32] Kernel_size1 = [(3,3),] Stride1 = [(1,1)]	(None, 50, 50, 32)	608	input_a[0][0]	
conv2d_2 (Conv2D)	Filter2= [32] Kernel_size2 = [(5,5),] Stride2 = [(1,1)]	(None, 50, 50, 32)	1632	input_a[0][0]	
conv2d_3 (Conv2D)	Filter3= [5] Kernel_size3 = [(7,7),] Stride3 = [(1,1)]	(None, 50, 50, 5)	7845	conv2d_1[0][0]	
conv2d_4 (Conv2D)	Filter4= [5] Kernel_size4 = [(7,7),] Stride4 = [(1,1)]	(None, 50, 50, 5)	7845	conv2d_2[0][0]	
concatenate_1 (Concatenate)		(None, 50, 50, 10)	0	conv2d_3[0][0] conv2d_4[0][0]	
dropout_1 (Dropout)		(None, 50, 50, 10)	0	concatenate_1[0][0]	
flatten_1 (Flatten)		(None, 25,000)	0	dropout_1[0][0]	
dense_1 (Dense)	Unit= [32]	(None, 32)	800,032	flatten_1[0][0]	
dense_2 (Dense)		(None, 14)	462	dense_1[0][0]	

Table A9

Specific structure of proposed CNN-CBAM.

		Layer (type)	Number of parameters	Output Shape	Number of parameters	Connected to	Total parameters
	Input Layer	input_a (InputLayer)		(None, 50, 50, 2)	0		231,522
	Convolution 1	conv2d_1 (Conv2D)	Filter1 = [128] Kernel_size1= [(3,3)] Stride1 = [(1,1)]	(None, 48, 48, 128)	2432	input_a[0][0]	
		conv2d_2 (Conv2D)	Filter2= [64] Kernel_size2= [(5,5)] Stride2 = [(1,1),] channel_attention_reduction_ratio1= [0.5]	(None, 44, 44, 64)	204,864	conv2d_1[0][0]	
CBAM module 1	Channel attention	global_max_pooling2d_1 (GlobalMaxPooling2D)		(None, 64)	0	conv2d_2[0][0]	
		global_average_pooling2d_1 (GlobalAvgPool2D)		(None, 64)	0	conv2d_2[0][0]	
		reshape_1 (Reshape)		(None, 1, 1, 64)	0	global_max_pooling2d_1 [0][0]	
		reshape_2 (Reshape)		(None, 1, 1, 64)	0	global_average_pooling2d_1 [0][0]	
		dense_1 (Dense)		(None, 1, 1, 32)	2080	reshape_1[0][0]	
		dense_2 (Dense)		(None, 1, 1, 64)	2112	reshape_2[0][0]	
		reshape_3 (Reshape)		(None, 1, 1, 64)	0	dense_1[0][0]	
		reshape_4 (Reshape)		(None, 1, 1, 64)	0	dense_1[0, 1]	
		add_1 (Add)		(None, 1, 1, 64)	0	dense_2[0][0]	
		activation_1 (Activation)		(None, 1, 1, 64)	0	reshape_3[0][0]	
		multiply_1 (Multiply)		(None, 44, 44, 64)	0	reshape_4[0][0]	
Spatial attention		lambda_1 (Lambda: max)		(None, 44, 44, 1)	0	activation_1[0][0]	
		lambda_2 (Lambda: mean)		(None, 44, 44, 1)	0	conv2d_2[0][0]	
		concatenate_1 (Concatenate)		(None, 44, 44, 2)	0	multiply_1[0][0]	
		conv2d_s1 (Conv2D)	spatial_attention kernel_size1= [(3,3)]	(None, 44, 44, 1)	18	lambda_1[0][0]	
		multiply_2 (Multiply)		(None, 44, 44, 64)	0	lambda_2[0][0]	
		add_2 (Add)		(None, 44, 44, 64)	0	concatenate_1[0][0]	
Convolution 2		conv2d_3 (Conv2D)	Filter3= [32] Kernel_size3= [(3,3)] Stride3 = [(1,1),]	(None, 42, 42, 32)	18,464	multiply_2[0][0]	
CBAM module 2	Channel attention	global_max_pooling2d_2 (GlobalMaxPooling2D)	channel_attention_reduction_ratio2= [0.5]	(None, 32)	0	conv2d_2[0][0]	
		global_average_pooling2d_2 (GlobalAvgPool2D)		(None, 32)	0	conv2d_3[0][0]	
		reshape_5 (Reshape)		(None, 1, 1, 32)	0	global_max_pooling2d_2 [0][0]	
		reshape_6 (Reshape)		(None, 1, 1, 32)	0	global_average_pooling2d_2 [0][0]	
		dense_3 (Dense)		(None, 1, 1, 16)	528	reshape_5[0][0]	
		dense_4 (Dense)		(None, 1, 1, 32)	544	reshape_6[0][0]	
		reshape_7 (Reshape)		(None, 1, 1, 32)	0	dense_3[0][0]	
		reshape_8 (Reshape)		(None, 1, 1, 32)	0	dense_3[0, 1]	
		add_3 (Add)		(None, 1, 1, 32)	0	dense_4[0][0]	
		activation_2 (Activation)		(None, 1, 1, 32)	0	dense_4[0, 1]	
		multiply_3 (Multiply)		(None, 42, 42, 32)	0	activation_2[0][0]	
						conv2d_3[0][0]	

(continued on next page)

Table A9 (continued)

	Layer (type)	Number of parameters	Output Shape	Number of parameters	Connected to	Total parameters
Spatial attention	lambda_3 (Lambda: max)		(None, 42, 42, 1)	0	multiply_3[0][0]	
	lambda_4 (Lambda: mean)		(None, 42, 42, 1)	0	multiply_3[0][0]	
	concatenate_2 (Concatenate)		(None, 42, 42, 2)	0	lambda_3[0][0] lambda_4[0][0]	
	conv2d_s2 (Conv2D)	spatial_attention kernel_size2= [(3,3)]	(None, 42, 42, 1)	18	concatenate_2[0][0]	
	multiply_4 (Multiply)		(None, 42, 42, 32)	0	multiply_3[0][0] conv2d_s2[0][0]	
	add_4 (Add)		(None, 42, 42, 32)	0	multiply_4[0][0] conv2d_3[0][0] add_4[0][0]	
GlobalAvgPool	global_average_pooling2d_3 (GlobalAvgPool2D)		(None, 32)	0		
Output	dense_5 (Dense)		(None, 14)	462	global_average_pooling2d_3 [0][0]	

References

- [1] Citaristi I. International energy agency—IEA. The Europa directory of international organizations 2022. Routledge; 2022. p. 701–2.
- [2] Zhao Y, De Palma J-F, Mosesian J, Lyons R, Lehman B. Line–line fault analysis and protection challenges in solar photovoltaic arrays. *IEEE Trans Industr Electron* 2012;60:3784–95.
- [3] Livera A, Theristis M, Makrides G, Georgiou GE. Recent advances in failure diagnosis techniques based on performance data analysis for grid-connected photovoltaic systems. *Renew Energy* 2019;133:126–43.
- [4] Momeni H, Sadooghi N, Farrokhisfar M, Gharibeh HF. Fault diagnosis in photovoltaic arrays using GBSSL method and proposing a fault correction system. *IEEE Trans Industr Inform* 2019;16:5300–8.
- [5] Pillai DS, Blaabjerg F, Rajasekar N. A comparative evaluation of advanced fault detection approaches for PV systems. *IEEE J Photovolt* 2019;9:513–27.
- [6] Li C, Yang Y, Zhang K, Zhu C, Wei H. A fast MPPT-based anomaly detection and accurate fault diagnosis technique for PV arrays. *Energ Conver Manage* 2021;234: 113950.
- [7] Pillai DS, Rajasekar N. A comprehensive review on protection challenges and fault diagnosis in PV systems. *Renew Sustain Energy Rev* 2018;91:18–40.
- [8] Dida M, Boughali S, Bechki D, Bouguettaia H. Output power loss of crystalline silicon photovoltaic modules due to dust accumulation in Saharan environment. *Renew Sustain Energy Rev* 2020;124:109787.
- [9] Kazem HA, Chaichan MT, Al-Waeli AH, Sopian K. A review of dust accumulation and cleaning methods for solar photovoltaic systems. *J Clean Prod* 2020;276: 123187.
- [10] Song Z, Liu J, Yang H. Air pollution and soiling implications for solar photovoltaic power generation: A comprehensive review. *Appl Energy* 2021;298:117247.
- [11] Salamat T, Ramahi A, Alamari K, Juaidi A, Abdallah R, Abdelkareem MA, et al. Effect of dust and methods of cleaning on the performance of solar PV module for different climate regions: comprehensive review. *Sci Total Environ* 2022;154050.
- [12] Younis A, Alhor Y. Modeling of dust soiling effects on solar photovoltaic performance: A review. *Solar Energy* 2021;220:1074–88.
- [13] Van Gompel J, Spina D, Develder C. Satellite based fault diagnosis of photovoltaic systems using recurrent neural networks. *Appl Energy* 2022;305:117874.
- [14] Araneo R, Lammens S, Grossi M, Bertone S. EMC issues in high-power grid-connected photovoltaic plants. *IEEE Trans Electromagn Compat* 2009;51:639–48.
- [15] Takashima T, Yamaguchi J, Otani K, Oozeki T, Kato K, Ishida M. Experimental studies of fault location in PV module strings. *Solar Energy Mater Solar Cells* 2009; 93:1079–82.
- [16] Roy S, Alam MK, Khan F, Johnson J, Flicker J. An irradiance-independent, robust ground-fault detection scheme for PV arrays based on spread spectrum time-domain reflectometry (SSTDR). *IEEE Trans Power Electron* 2017;33:7046–57.
- [17] Madeti SR, Singh S. A comprehensive study on different types of faults and detection techniques for solar photovoltaic system. *Solar Energy* 2017;158:161–85.
- [18] Chen S-Q, Yang G-J, Gao W, Guo M-F. Photovoltaic fault diagnosis via semisupervised ladder network with string voltage and current measures. *IEEE J Photovolt* 2020;11:219–31.
- [19] Chen Z, Han F, Wu L, Yu J, Cheng S, Lin P, et al. Random forest based intelligent fault diagnosis for PV arrays using array voltage and string currents. *Energ Conver Manage* 2018;178:250–64.
- [20] Lu X, Lin P, Cheng S, Lin Y, Chen Z, Wu L, et al. Fault diagnosis for photovoltaic array based on convolutional neural network and electrical time series graph. *Energ Conver Manage* 2019;196:950–65.
- [21] Abd El-Ghany HA, Elgebaly AE, Taha IBM. A new monitoring technique for fault detection and classification in PV systems based on rate of change of voltage-current trajectory. *Int J Electric Power Energy Syst* 2021;133:107248.
- [22] Liu Y, Ding K, Zhang J, Li Y, Yang Z, Zheng W, et al. Fault diagnosis approach for photovoltaic array based on the stacked auto-encoder and clustering with I-V curves. *Energ Conver Manage* 2021;245:114603.
- [23] Pan T, Chen J, Xie J, Chang Y, Zhou Z. Intelligent fault identification for industrial automation system via multi-scale convolutional generative adversarial network with partially labeled samples. *ISA Trans* 2020;101:379–89.
- [24] Ishaque K, Salam Z. An improved modeling method to determine the model parameters of photovoltaic (PV) modules using differential evolution (DE). *Solar Energy* 2011;85:2349–59.
- [25] Ma M, Zhang Z, Yun P, Xie Z, Wang H, Ma W. Photovoltaic module current mismatch fault diagnosis based on IV data. *IEEE J Photovolt* 2021;11:779–88.
- [26] Zhang Z, Ma M, Wang H, Wang H, Ma W, Zhang X. A fault diagnosis method for photovoltaic module current mismatch based on numerical analysis and statistics. *Solar Energy* 2021;225:221–36.
- [27] Drews A, De Keizer A, Beyer HG, Lorenz E, Betcke J, Van Sark W, et al. Monitoring and remote failure detection of grid-connected PV systems based on satellite observations. *Solar Energy* 2007;81:548–64.
- [28] Garoudja E, Harrou F, Sun Y, Kara K, Chouder A, Silvestre S. Statistical fault detection in photovoltaic systems. *Solar Energy* 2017;150:485–99.
- [29] Mellit A, Tina GM, Kalogirou SA. Fault detection and diagnosis methods for photovoltaic systems: A review. *Renew Sustain Energy Rev* 2018;91:1–17.
- [30] Triki-Lahiani A, Abdelghani AB-B, Slama-Belkhodja I. Fault detection and monitoring systems for photovoltaic installations: A review. *Renew Sustain Energy Rev* 2018;82:2680–92.
- [31] Spataru S, Sera D, Kerekes T, Teodorescu R. Monitoring and fault detection in photovoltaic systems based on inverter measured string IV curves. In: 31st European Photovoltaic Solar Energy Conference and Exhibition. WIP Wirtschaft und Infrastruktur GmbH & Co Planungs KG; 2015. p. 1667–74.
- [32] Chen Z, Chen Y, Wu L, Cheng S, Lin P. Deep residual network based fault detection and diagnosis of photovoltaic arrays using current-voltage curves and ambient conditions. *Energ Conver Manage* 2019;198:111793.
- [33] Gao W, Wai R-J. A novel fault identification method for photovoltaic array via convolutional neural network and residual gated recurrent unit. *IEEE Access* 2020; 8:159493–510.
- [34] Chine W, Mellit A, Lughiv A, Malek A, Sulligoi G, Pavan AM. A novel fault diagnosis technique for photovoltaic systems based on artificial neural networks. *Renew Energy* 2016;90:501–12.
- [35] Li B, Delpha C, Migan-Dubois A, Diallo D. Fault diagnosis of photovoltaic panels using full I-V characteristics and machine learning techniques. *Energ Conver Manage* 2021;248:114785.
- [36] Fadhel S, Delpha C, Diallo D, Bahri I, Migan A, Trabelsi M, et al. PV shading fault detection and classification based on IV curve using principal component analysis: application to isolated PV system. *Solar Energy* 2019;179:1–10.
- [37] Liu Y, Ding K, Zhang J, Lin Y, Yang Z, Chen X, et al. Intelligent fault diagnosis of photovoltaic array based on variable predictive models and I–V curves. *Solar Energy* 2022;237:340–51.
- [38] Singh R, Sharma M, Rawat R, Banerjee C. An assessment of series resistance estimation techniques for different silicon based SPV modules. *Renew Sustain Energy Rev* 2018;98:199–216.
- [39] Li Y, Ding K, Zhang J, Chen F, Chen X, Wu J. A fault diagnosis method for photovoltaic arrays based on fault parameters identification. *Renew Energy* 2019; 143:52–63.
- [40] Chen Z, Wu L, Cheng S, Lin P, Wu Y, Lin W. Intelligent fault diagnosis of photovoltaic arrays based on optimized kernel extreme learning machine and I-V characteristics. *Appl Energy* 2017;204:912–31.
- [41] Bressan M, El Basri Y, Galeano AG, Alonso C. A shadow fault detection method based on the standard error analysis of IV curves. *Renew Energy* 2016;99:1181–90.

- [42] Ma M, Zhang Z, Xie Z, Yun P, Zhang X, Li F. Fault diagnosis of cracks in crystalline silicon photovoltaic modules through IV curve. *Microelectr Reliab* 2020;114: 113848.
- [43] Lin P, Qian Z, Lu X, Lin Y, Lai Y, Cheng S, et al. Compound fault diagnosis model for photovoltaic array using multi-scale SE-ResNet. *Sustain Energy Technol Assessm* 2022;50:101785.
- [44] Huang J-M, Wai R-J, Yang G-J. Design of hybrid artificial bee colony algorithm and semi-supervised extreme learning machine for PV fault diagnoses by considering dust impact. *IEEE Trans Power Electron* 2019;35:7086–99.
- [45] Li B, Migan-Dubois A, Delpha C, Diallo D. Evaluation and improvement of IEC 60891 correction methods for I-V curves of defective photovoltaic panels. *Solar Energy* 2021;216:225–37.
- [46] Schill C, Brachmann S, Koehl M. Impact of soiling on IV-curves and efficiency of PV-modules. *Solar Energy* 2015;112:259–62.
- [47] Chanchangi YN, Ghosh A, Sundaram S, Mallick TK. Dust and PV performance in Nigeria: A review. *Renew Sustain Energy Rev* 2020;121:109704.
- [48] Silvestre S, Boronat A, Clouder A. Study of bypass diodes configuration on PV modules. *Appl Energy* 2009;86:1632–40.
- [49] Heinrich M, Meunier S, Same A, Queval L, Darga A, Oukhellou L, et al. Detection of cleaning interventions on photovoltaic modules with machine learning. *Appl Energy* 2020;263:114642.
- [50] Al-Addous M, Dalala Z, Alawneh F, Class CB. Modeling and quantifying dust accumulation impact on PV module performance. *Solar Energy* 2019;194:86–102.
- [51] Hachicha AA, Al-Sawafta I, Said Z. Impact of dust on the performance of solar photovoltaic (PV) systems under United Arab Emirates weather conditions. *Renew Energy* 2019;141:287–97.
- [52] Soon JJ, Low K-S. Photovoltaic model identification using particle swarm optimization with inverse barrier constraint. *IEEE Trans Power Electron* 2012;27: 3975–83.
- [53] Bastidas-Rodríguez JD, Franco E, Petrone G, Ramos-Paja CA, Spagnuolo G. Model-based degradation analysis of photovoltaic modules through series resistance estimation. *IEEE Trans Industr Electron* 2015;62:7256–65.
- [54] Commission I E. IEC 60891, photovoltaic devices. Procedures for temperature and irradiance corrections to measure IV characteristics. In: International electrotechnical commission: Geneva, Switzerland; 2007.
- [55] Bauomy MF, Gamal H, Shaltout AA. Solar PV DC nanogrid dynamic modeling applying the polynomial computational method for MPPT. *Adv Clean Energy Technol Elsevier* 2021;19–87.
- [56] Lappalainen K, Valkealahti S. Experimental study of the maximum power point characteristics of partially shaded photovoltaic strings. *Appl Energy* 2021;301: 117436.
- [57] Chen X, Ding K, Zhang J, Yang Z, Liu Y, Yang H. A two-stage method for model parameter identification based on the maximum power matching and improved flow direction algorithm. *Energ Conver Manage* 2023;278:116712.
- [58] Wang Z, Oates T. Imaging time-series to improve classification and imputation. In: Twenty-Fourth International Joint Conference on Artificial Intelligence; 2015.
- [59] Villalva MG, Gazoli JR, Ruppert Filho E. Comprehensive approach to modeling and simulation of photovoltaic arrays. *IEEE Trans Power Electron* 2009;24:1198–208.
- [60] Villalva MG, Gazoli JR, Ruppert Filho E. Modeling and circuit-based simulation of photovoltaic arrays. In: 2009 Brazilian Power Electronics Conference. IEEE; 2009. p. 1244–54.
- [61] De Soto W, Klein SA, Beckman WA. Improvement and validation of a model for photovoltaic array performance. *Solar Energy* 2006;80:78–88.
- [62] Alzubaidi I, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J Big Data* 2021;8:1–74.
- [63] Woo S, Park J, Lee J-Y, Kweon IS. Cbam: Convolutional block attention module. In: Proceedings of the European Conference on Computer Vision (ECCV); 2018. p. 3–19.
- [64] Qu J, Qian Z, Pei Y, Wei L, Zareipour H, Sun Q. An unsupervised hourly weather status pattern recognition and blending fitting model for PV system fault detection. *App Energ* 2022;319:119271.
- [65] Hong Y-Y, Pula RA. Detection and classification of faults in photovoltaic arrays using a 3D convolutional neural network. *Energy* 2022;246:123391.
- [66] Eskandari A, Milimonfared J, Aghaei M. Line-line fault detection and classification for photovoltaic systems using ensemble learning model based on IV characteristics. *Solar Energy* 2020;211:354–65.
- [67] Wang Y, Li J. Credible intervals for precision and recall based on a K-fold cross-validated beta distribution. *Neural Comput* 2016;28:1694–722.
- [68] Hossin M, Sulaiman MN. A review on evaluation metrics for data classification evaluations. *Int J Data Min Knowledge Manag Process* 2015;5:1.
- [69] Thiel M, Romano MC, Kurths J. How much information is contained in a recurrence plot? *Phys Lett A* 2004;330:343–9.
- [70] Garcia S, Herrera F. An extension on“ statistical comparisons of classifiers over multiple data sets” for all pairwise comparisons. *J Mach Learn Res* 2008;9.
- [71] Gibbons JD, Fielden JDG. Nonparametric Statistics: An Introduction. Sage; 1993.
- [72] Luengo J, García S, Herrera F. A study on the use of statistical tests for experimentation with neural networks: analysis of parametric test conditions and non-parametric tests. *Exp Syst Appl* 2009;36:7798–808.
- [73] Zimmerman DW, Zumbo BD. Relative power of the Wilcoxon test, the Friedman test, and repeated-measures ANOVA on ranks. *J Exp Educ* 1993;62:75–86.