

# An efficient approach for diagnosing faults in photovoltaic array using 1D-CNN and feature selection Techniques

Yousif Mahmoud Ali <sup>a</sup>, Lei Ding <sup>a,\*</sup> , Shiyao Qin <sup>b,\*</sup>

<sup>a</sup> Key Laboratory of Power System Intelligent Dispatch and Control of Ministry of Education, School of Electrical Engineering, Shandong University, Jinan 250061, China

<sup>b</sup> China Electric Power Research Institute, 15 Xiaoying East Road, Qinghe, Beijing, 100192, China



## ARTICLE INFO

### Keywords:

Photovoltaic array  
Current-voltage curve  
Fault diagnosis  
Convolutional Neural Network  
Feature selection

## ABSTRACT

Diagnosing faults in Photovoltaic (PV) systems is essential for operation and maintenance. Selecting relevant features is necessary for successful fault diagnosis because redundant and irrelevant features reduce fault diagnosing accuracy. This paper proposes a novel and efficient approach to diagnosing faults in PV systems. The Feature Selection and Fault Diagnosis (FSFD) method is executed for diagnosing five types of faults in PV array (PVA): partial shading condition, line-line fault, arc fault, open-circuit fault, and degradation fault. Firstly, a PVA modeling method using MATLAB/Simulink is employed to simulate I-V curves and extract their features. Next, a feature permutation technique-based method is proposed for selecting the most relevant features. A simple and accurate one-dimensional convolutional neural network (1D-CNN) model is developed to classify the faults based on the selected features. Finally, a confusion matrix is utilized to evaluate the performance of the trained model. Three datasets of PVAs have been utilized to evaluate the effectiveness of the proposed FSFD method. The results indicate that the FSFD method has effectively identified the best five features out of eight for training the 1D-CNN model. The trained model has achieved diagnosing accuracy rates of 99.85%, 99.73%, and 99.97% in series-parallel PVA, total cross-tied PVA, and series PVA datasets, respectively. The proposed method accurately diagnoses single faults in three PVA configurations. Therefore, we recommend conducting additional studies to improve the proposed method for diagnosing hybrid faults.

## 1. Introduction

The rapid expansion of photovoltaic (PV) power generation, particularly in the distribution sector, has led to a substantial rise in energy output [1]. However, PV systems are prone to multiple faults, such as arc faults [2], line ground, and line-line faults, which damage PV modules and cables [3]. These faults may reduce power production and pose potential safety hazards, such as fire [4]. However, according to NEC article 690, PV arrays are protected against line-to-line (LLF), line-to-ground (LGF), and arc faults (AF) using overcurrent protection devices (OCPDs), ground fault detection and interruption (GFDI), and arc fault circuit interrupters (AFCIs) [5]. Nevertheless, the devices fail to identify the faults in the presence of the blocking diodes and maximum power point tracker (MPPT) technique [6,7]. Several methods are suggested for detecting and classifying a wide range of electrical faults in PV systems to address the limitations of conventional protection devices. These methods can be classified into methods based on analytical models and methods based on AI models.

Firstly, the category of methods based on analytical models includes methods that use mathematical analysis and comparison concepts for detecting the faults in the PV system. Authors in [8] proposed a method to monitor the current deviations from average values and compare them to a pre-determined threshold to detect and classify the faults in PV systems. In [9], the rate of changes in voltages and currents at PV strings and the PV array's current are used for detecting and discriminating the faulty conditions based on predefined thresholds. A real-time fault detection system based on statistical analysis is proposed in [7]. The faults are detected when Z-Scores exceed defined thresholds by calculating the Z-Score for each string based on actual string power versus the simulated mean. This method enables the detection of faults like PV module failures, shading, and full-string failures. Besides, the statistical *t*-test approach is used in [10] to compare the real power from the output PV system to the reference values for detecting the faults in a PV system.

Even though the analytical model-based methods have a simple model and efficient computation, they have some limitations. For example, the fault detection algorithm is sensitive to the predefined

\* Corresponding author.

E-mail addresses: [dinglei@sdu.edu.cn](mailto:dinglei@sdu.edu.cn) (L. Ding), [qinshy@epri.sgcc.com.cn](mailto:qinshy@epri.sgcc.com.cn) (S. Qin).

threshold, and estimating ideal thresholds to detect and discriminate the PV faults is difficult. Moreover, these methods used several thresholds to classify multiple faults, which increases model complexity and computation cost.

Secondly, AI model-based methods aid in overcoming the thresholding challenges and accurately detect and classify multiple faults due to their high capability in data classification [1,6]. In this regard, machine learning (ML) and Deep learning (DL) algorithms are employed in recent research to identify and diagnose faults in PV systems. The algorithms of K-Nearest Neighbors (KNN) [4,11], Support Vector Machine (SVM) [6], Random Under-Sampling Boosting (RUSBoost) [12], Probabilistic Neural Network (PNN) [13,14], Artificial Neural Network (ANN) [15–21], Convolutional Neural Networks (CNN) [22–24], and Variable Prediction Model (VPM) [25] have utilized as a classifier to enhance fault detection methods. Generally, the AI model-based method relies on learning the model to discover patterns and correlations within PV system data. Moreover, the effectiveness of the learned model depends on the quality of the learning data (input data). In recent research, various features have been proposed to be used to learn the AI model and achieve high fault detection accuracy. In this study, we can classify AI methods based on the type of features into time series-based methods and current–voltage (I-V) characteristics-based methods.

The time series-based methods have used the power sequences as input data to detect and diagnose the PV system faults. A study [17] has proposed a monitoring approach to detect the disconnection of modules in a PV array using the variations of the power and irradiation. Several forms of data are used as input data to learn the intelligent algorithms for diagnosing LLF and OCF. For instance, the voltage and current are used in [6,20,22]. Irradiation and temperature were added to the voltage and current to enhance the dataset in the study [11]. Furthermore, the PV array output power was utilized in [23]. In [13] and [15], the researchers have employed the voltages and currents of PV strings, along with irradiation and temperature, as input data for diagnosing LLF and open-circuit fault (OCF), partial shading condition (PSC) and degradation fault (DF).

The methods based on I-V characteristics have employed features of the I-V curve, such as short circuit current ( $I_{sc}$ ), open circuit voltage ( $V_{oc}$ ), current at maximum power point ( $I_{mp}$ ), voltage at maximum power point ( $V_{mp}$ ), maximum power ( $P_{mp}$ ), and fill factor (FF), as input data to enhance the diagnosis of faults in photovoltaic systems, due to their higher sensitivity to these faults. For instance, the features of  $I_{mp}$ ,  $V_{mp}$ , and  $P_{mp}$  are used as input data to learn the model in [25] to detect SCF, PSC, and DF. Moreover, the intelligent model in [4] has used  $I_{mp}$ ,  $V_{mp}$ , and  $P_{mp}$  to detect LLF, OCF, and PSC. In [12,14,18,21],  $I_{mp}$ ,  $V_{mp}$ , and  $P_{mp}$ , along with irradiation and temperature, have been selected as input data to train their models for diagnosing LLF, OCF, DF, and PSC in [12], and LLF, OCF, and PSC in [18], as well as LLF and OCF in [14] and [21]. The researchers in [16] have utilized eight input data, including  $I_{sc}$ ,  $V_{oc}$ ,  $I_{mp}$ , and  $V_{mp}$ , in addition to PV array voltage and current, irradiation, and temperature to diagnose four types of PV array faults, including LLF, OCF, PSC, and DF.

It is noteworthy that the reviewed AI model-based methods have the following limitations:

1. Time series-based methods have a primary constraint because their features are insufficient for accurately diagnosing various PV faults. As a result, these methods have focused on diagnosing particular faults, such as OCF in [17], arc faults in [24], PSC and faulty modules in [20], and LLF and OCF in [6,11,13,22,23].
2. Time series features alone are inadequate for effectively differentiating between PV array faults. For instance, PSC and LLF exhibit similar voltage, current, and power characteristics. Therefore, the diagnosis models in [15] and [20] could not distinguish PSC from other faults, resulting in low accuracies for diagnosing these faults of 77.36 % and 89.9 %, respectively.

3. Using irrelevant features in the method based on I-V characteristics has resulted in the inability of the detection model in [4] to differentiate between PSC and LLF and reduced diagnosis accuracy of four faults to 93.4 % in [16].
4. The studies in [4,14,21] have only concentrated on using I-V characteristics to diagnose two faults in PVAs.
5. The impact of low irradiation and high impedance on the I-V curve characteristics is crucial for the precise detection of LLFs. This constraint was not considered in studies [4,12,16,18,25], resulting in misclassification and reduced diagnostic accuracy of LLF (. e.g., 80 % in [4]).

In this paper, we proposed an efficient approach for diagnosing faults in PV arrays using an AI model and feature selection technique to overcome the limitations of thresholding and learning data quality. Among AI algorithms, the CNN algorithms have received significant attention in recent research due to their superior performance and data classification abilities, which can address the limitation of estimating accurate thresholds in conventional analytical methods. CNN algorithms have been successfully used to detect PV system faults [22,23]. Therefore, this paper proposes using a CNN algorithm to develop the fault classification model.

The main steps of the proposed approach include. 1) collecting the dataset. For this step, the basic features are collected from the I-V curve characteristics of PV arrays under normal and faulty scenarios. Then, 2) the permutation importance algorithm as a feature selection approach is used to find the best features to enhance the data quality and obtain greater efficiency and accuracy in the classification procedure. Finally, 3) One dimension CNN (1D-CNN) model is developed for fault detection.

The proposed 1D-CNN algorithm consists of a convolutional layer, a Maxpooling layer, two fully connected layers, and a softmax activation layer. In order to adapt training of the 1D-CNN model on the datasets, various combinations of the hyperparameter values are tested, and the most successful combination is chosen. The results demonstrate that the proposed method accurately detects and diagnoses PV faults, even in noisy data and variations in PVA configuration scales. Furthermore, the results confirm the proposed method's reliability and accuracy on different PVA configurations. The primary contributions of this study might be encapsulated as follows:

- 1) A novel Feature Selection and Fault Diagnosis (FSFD) method for PVA has been developed.
- 2) This study has proposed different datasets under variable operating conditions, an efficient classification model, and a feature selection method to address the challenges of fault detection and diagnosis in PVA.
- 3) The permutation importance algorithm has been used to enhance learning data quality, improve classification accuracy for the learning model, and minimize training dataset size.
- 4) The performance of the proposed model has been evaluated on different configurations and scales of PVAs, and up to our knowledge, no previous study has investigated the evaluation of their AI model in different configurations and scales of PVAs.
- 5) The proposed method accurately detects and diagnoses five faults, including PSC, LLF, AF, OCF, and DF in different PVA configurations.

The subsequent sections of this work are structured as follows: Section 2 presents an analysis of PVA performance under various faults. Section 3 provides an in-depth explanation of the suggested methodology. Section 4 validates the efficacy of the proposed fault diagnosis method through simulation data for three PVA configurations and compares it with other methods. Lastly, section 5 summarizes the conclusions of the study.

## 2. Analysis of PVA performance under faults

This section explores the influence of the faults in the PV system. To do this, we employed 20 PV modules to form a 5 (series) x 4 (parallel) PVA, which has been modeled in MATLAB/Simulink based on the emulation circuits reported in [26] and [27] to assess the impact of faults on the I-V curve features. The parameters for the PV module under the standard test condition (STC) are concisely outlined in Table 1. The subsequent sections assess the performance of the PVA based on I-V curve characteristics for different fault scenarios, including PSC, LLF, AF, OCF, DF, and LGF, compared to its normal state under standard test conditions (STC).

### 2.1. PSC

PSC occurs within a PV system when an object or shadow covers a fraction of the PV modules, as depicted in Fig. 1(F1). PSC leads to a decrease in the power generation of the PVA and the formation of hot-spots within the shaded panel, accelerating its aging process [1]. To demonstrate power loss under PSCs, the irradiation for three PV modules in a single string is reduced to low levels, such as 800w/m<sup>2</sup> and 300w/m<sup>2</sup>, while keeping the other PV modules at the standard irradiation of 1000 w/m<sup>2</sup>. The analysis of the I-V curves in Fig. 2 indicates a drop in Pmp and Imp, with just a slight increase in Vmp. However, Isc and Voc remain unchanged. It is observed that the decrease in Pmp and Isc is directly related to the level of irradiation on the shaded modules, as shown in Fig. 2.

### 2.2. LLF

LLF arises from an accidental connection between two conductors carrying electric current in a PVA. LLFs can cause damage to the PV modules, the wiring, and the inverter, as well as reduce the output power and efficiency of the PVA. Also, it can increase the risk of fire hazards and electrical shock [1,26]. Typically, an unnoticed LLF fault leads to a disastrous alteration in the I-V characteristics of the PVA compared to its normal state. Fig. 3 demonstrates the I-V curve characteristics for LLF in two PV modules inside a single string, as depicted by Fig. 1(F2). It is evident from Fig. 3 that the Pmp and Imp are decreased due to LLF. While Vmp, Isc, and Voc remain unaffected. In addition, the decrease in fault impedance and the increase in faulty strings directly result in a proportional increase in the current loss, as seen in curves 2 and 3 of Fig. 3.

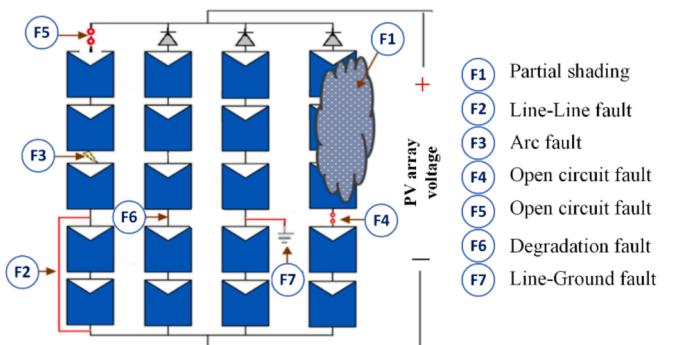
### 2.3. LGF

LGF is a short-circuit fault that occurs when an accidental connection between conductors carrying electric current and earth-ground conductors in a grounded PV system [1,18,28]. Similar to the LLF fault, the LGF fault can decrease the output power and efficiency of the PVA. Additionally, it can elevate the likelihood of fire dangers and electrical shock [1]. In order to assess the effect of LGF on the decrease in power production, the short circuit between two PV modules and the ground inside the PV string is analyzed, as depicted in Fig. 1 (F7). The analysis of the I-V curve characteristics shown in Fig. 4 indicates that the LGF undergoes changes that are comparable to those of the LLF, as

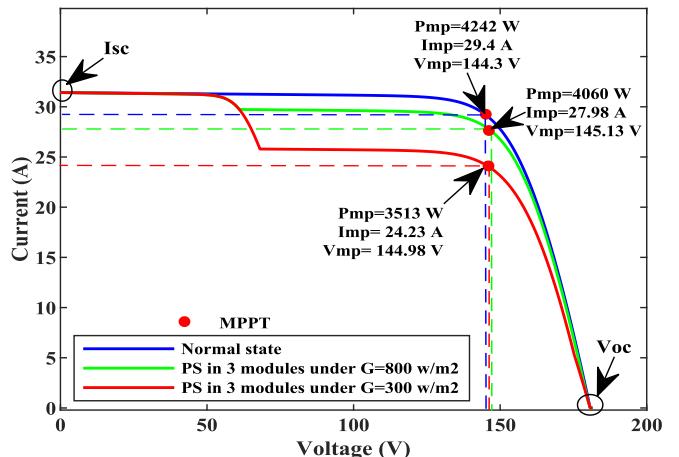
**Table 1**

The standard specifications of the PV module.

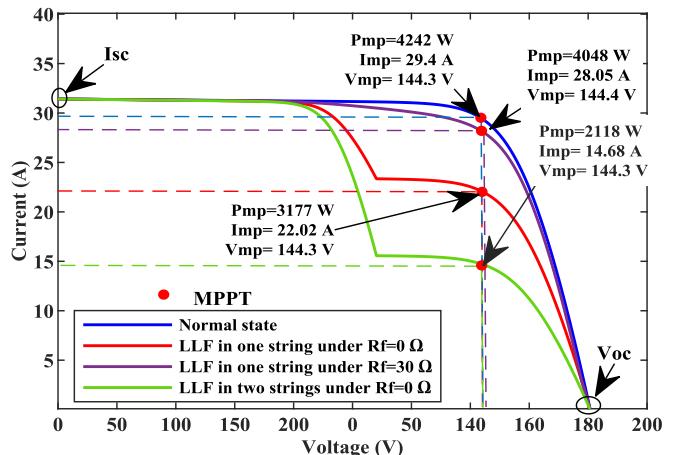
Parameters	Value
Maximum power $P_{mp}$	213.15
The voltage at maximum power point $V_{mp}$	29
The current at maximum power point $I_{mp}$	7.35
Short-circuit current $I_{sc}$	7.84
Open-circuit voltage $V_{oc}$	36.3



**Fig. 1.** The common PVA faults.



**Fig. 2.** I-V curve characteristics of PSC at STC.



**Fig. 3.** I-V curve characteristics of LLF under zero and non-zero fault impedance at STC.

demonstrated by curves 2 and 3 in Fig. 3. Therefore, in this study, we have considered LGF as a particular case of the LLFs.

### 2.4. AF

AF occurs when there is a poor connection at a joint in a conductor or when there is a failure in the insulation of a cable. An insecure connection can establish an electrical pathway through the atmosphere, initiating an electrical fire [2,26]. Typically, PVA has very low impedance in the connection between the modules [26]. AF results in a

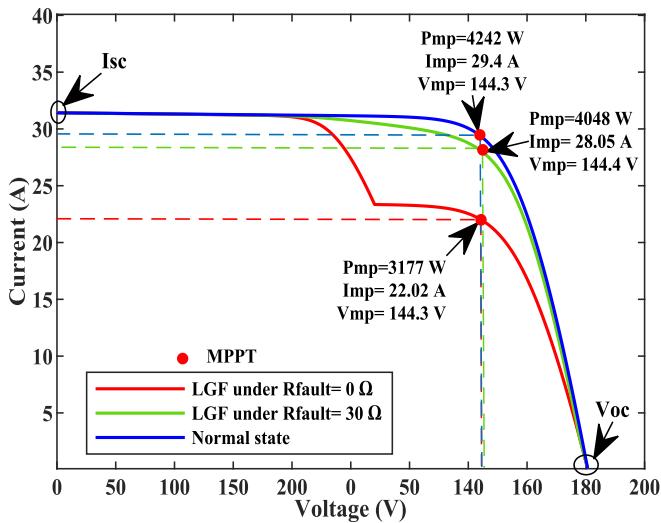


Fig. 4. I-V curve characteristics of LGF at STC.

decrease in the output power due to an increase in the series resistance between the connections. This study examines AF between PV modules, as shown in Fig. 1 (F3), using the methods described in [26] to recreate the AF with a high impedance in the simulation study. According to Fig. 5, when the AF impedance between two modules is large, there is a decrease in Isc, Imp, and Pmp. However, the reduction in Vmp is minor. However, Voc remains unchanged. Furthermore, there is a direct correlation between the decrease in Isc and Imp currents and the increase in arc impedance. As depicted in curve 4 of Fig. 5, when the impedance of a fault reaches a significantly high value (100 Ω), Isc and Imp decline to their lowest levels relative to the normal condition.

## 2.5. OCF

OCF occurs when there is a disconnection in a PV module or PV string, as depicted in Fig. 1 (F4) and (F5). These disconnections can be caused by insufficient soldering, breakage, blowing a string fuse, or disconnection of wires that link the components of the PV system. These faults result in a notable decrease in the system's output power [13,18,19]. Fig. 6 illustrates the influence of disconnection modules in the PVA. Upon analyzing the curves in Fig. 6, it is clear that disconnecting PV strings from the array leads to a substantial fall in Pmp, Isc, and Imp, whereas Vmp experiences a slight reduction compared to the normal state curve. Furthermore, the Voc remains unchanged. Furthermore, when disconnecting numerous strings (such as two strings), Pmp, Isc, and Imp values decrease sharply, as seen in curve 3 of Fig. 6.

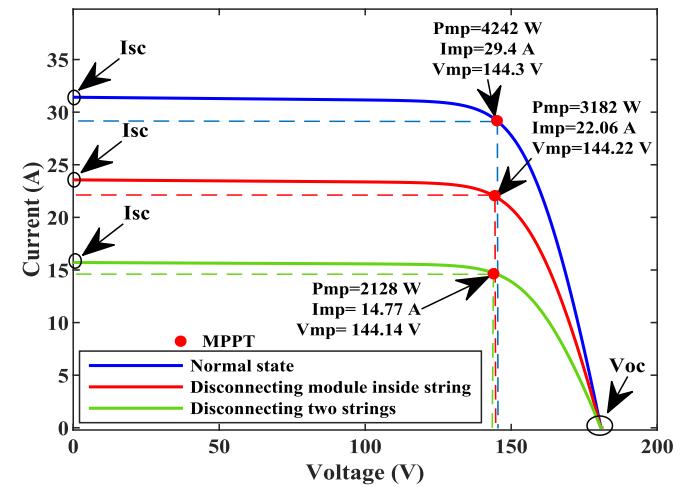


Fig. 6. I-V curve characteristics of OCF at STC.

Isc, and Imp, whereas Vmp experiences a slight reduction compared to the normal state curve. Furthermore, the Voc remains unchanged. Furthermore, when disconnecting numerous strings (such as two strings), Pmp, Isc, and Imp values decrease sharply, as seen in curve 3 of Fig. 6.

## 2.6. DF

DF in a PVA may occur due to increased series or decreased shunt resistance [12,15]. The increasing series resistance decreases the output voltage of the PV module and the entire string of cells [18]. Fig. 7 illustrates the characteristics of the I-V curve for the DF resulting from increased resistance between two PV modules within a single string, as illustrated in Fig. 1 (F6). Compared to the normal state, there is a decline in Pmp, Imp, and Vmp, along with a slight decrease in Isc, while Voc remains unchanged.

In addition to the characteristics of I-V curves, the fill factor (FF) plays an essential role in estimating the efficiency of PVA. The term FF refers to the proportion of the highest power output of the solar cell compared to the multiplication of its open-circuit voltage (Voc) and short-circuit current (Isc) [29]. The data presented in Table 2 demonstrates the I-V curve characteristics and FF for the investigated faults. Upon analysis, it is clear that the FF, Pmp, and Imp values decreased in

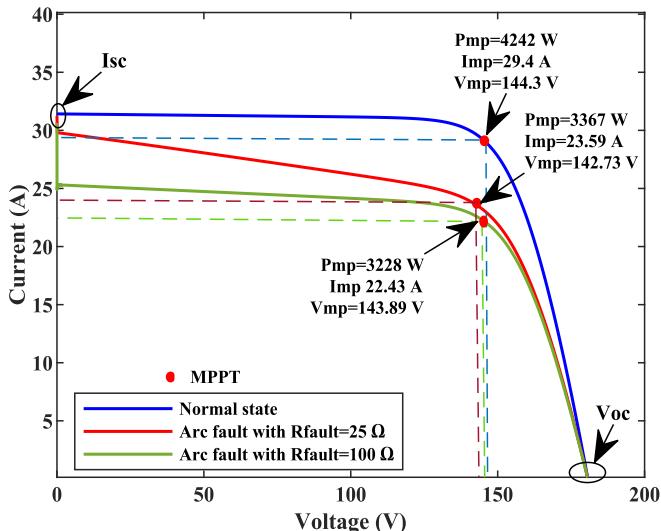


Fig. 5. I-V curve characteristics of AF at STC.

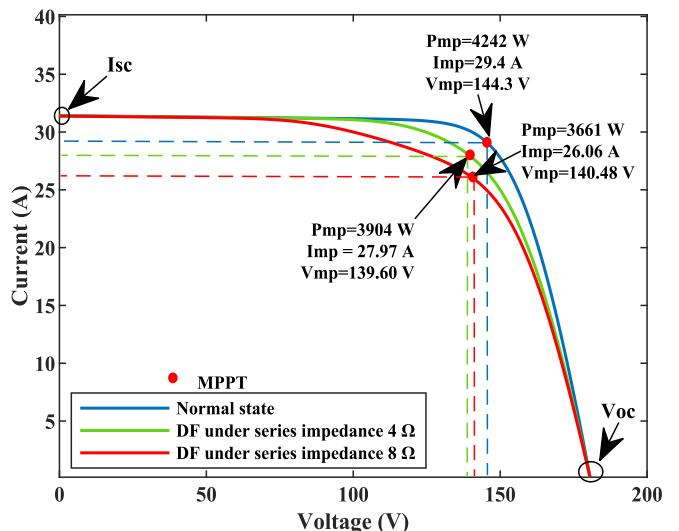


Fig. 7. I-V curve characteristics of DF at STC.

**Table 2**

Summary of SP-PVA characteristics operated in different fault conditions.

Operating state	Fault Condition	I-V curve characteristics					FF
		Pmp	Isc	Voc	Imp	Vmp	
Normal	—	4242	31.43	181.5	29.40	144.30	0.7436
PSC	G = 800 W/m <sup>2</sup>	4060	31.43	181.5	27.98	145.13	0.7117
	G = 300 W/m <sup>2</sup>	3513	31.43	181.5	24.23	144.98	0.6158
LLF	Fault impedance = 0 Ω	3177	31.43	181.5	22.02	144.30	0.5569
	Fault impedance = 30 Ω	4048	31.43	181.5	28.05	144.40	0.7096
AF	Arc impedance = 25 Ω	3367	31.20	181.5	23.59	142.73	0.5946
	Arc impedance = 100 Ω	3228	26.42	181.5	22.43	143.89	0.6732
OCF	Disconnect one string	3182	24.45	181.5	22.06	144.22	0.6387
	Disconnect two strings	2128	16.20	181.5	14.77	144.14	0.5000
DF	Series impedance = 4 Ω	3904	31.39	181.5	27.96	139.60	0.6852
	Series impedance = 8 Ω	3661	31.37	181.5	26.06	140.48	0.6430

all investigated faults. Furthermore, there are reductions in OCF, AF, and DF, accompanied by a minor rise in Vmp for PSC and LLF under high fault impedance, such as 30 Ω. In addition, the Isc experiences a significant decrease in the AF and OCF cases, with a slight drop in the DF, while it remains constant in the LLF and PSC cases. Significantly, the Voc value remains the same in all fault cases.

Finally, this analysis reveals distinct differences in the I-V curve characteristics for PVA faults compared to the normal state of the PVA. Furthermore, it has been observed that certain faults have analogous impacts on the I-V curve features. For instance, the similarity between PSC and LLF under high fault impedance (30 Ω) and the similarity between an AF and an OCF. This similarity presents a challenge in discriminating between these types of faults. In addition to the above investigation, the study [30] has demonstrated that the output voltages and currents of PVA are significantly influenced by temperature (T) and irradiation (G). Therefore, this study will consider T and G, I-V curve characteristics, and FF as diagnosis features to identify the optimum features for training the proposed DL model to diagnose PVA faults.

### 3. Proposed methodology

This section proposes a method based on an intelligent model for detecting and diagnosing faults in the PV system. Fig. 8 illustrates a block diagram depicting the framework of the developed diagnostic method. The proposed method includes five stages, which are 1) data acquisition, 2) data pre-processing, 3) feature selection, 4) develop the fault detection and diagnosis (FDD) model based on 1D-CNN, and 5) evaluate model performance. The details of these stages will be discussed comprehensively in the subsequent subsections.

### 3.1. Data acquisition

A single-diode model (SDM) is the most commonly used in the literature to model PV modules [12]. Here, the SDM indicated in Fig. 9 and Eq. (1) is used.

$$I = I_{ph} - I_0 \left( \frac{V + IR_s}{ns \times Vt} - 1 \right) - \left( \frac{V + IR_s}{R_{sh}} \right) \quad (1)$$

$$Vt = \frac{A \times K \times T}{q} \quad (2)$$

$$P = I \times V \quad (3)$$

Where I, V, and P are, respectively, the PV module's output current, voltage, and power,  $I_{ph}$  is the cell Photocurrent,  $I_0$  is the saturation current of the diode,  $R_s$  is series resistances,  $R_{sh}$  is shunt resistances respectively, ns is the total number of cells of the series,  $Vt$  is the thermal voltage calculated as shown in Eq. (2). Where A is the ideal factor of

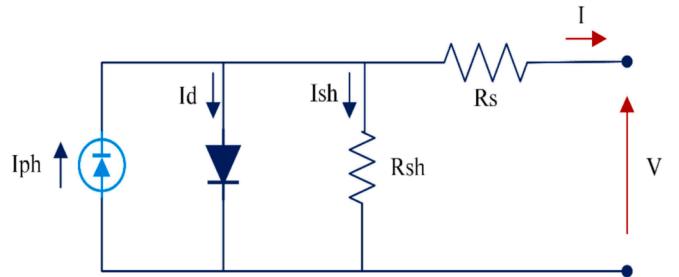


Fig. 9. Single diode model of PV cell.

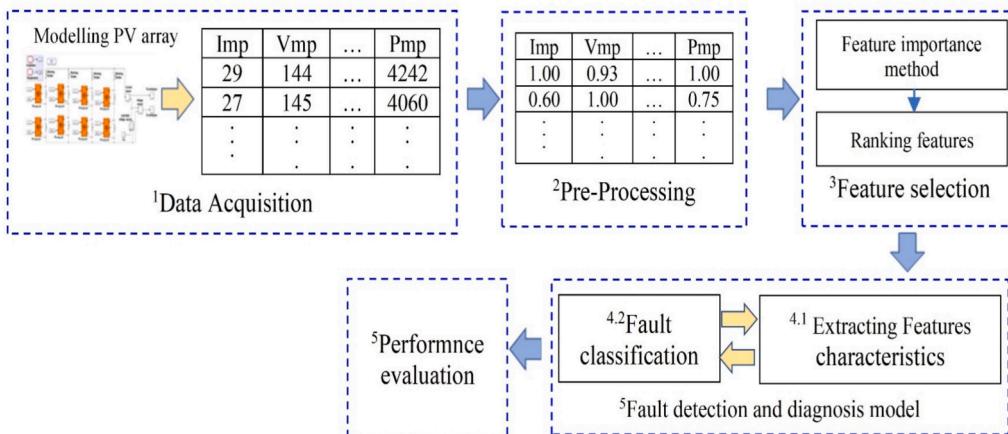


Fig. 8. Framework of the proposed fault diagnosis method.

diode, K is the Boltzmann's constant, T is the module operating temperature in kelvin degree, and q is the electronic charge.

A PVA modeling method using MATLAB/Simulink based on the I-V curve test circuits described in [26] and [27] is utilized to collect the dataset needed for learning the AI model. According to the modeling setup, the short circuit that simulates LLF is effectively regulated by a set of tiny and high resistors. At the same time, the OCF is simulated by inserting an exceedingly high resistor between the adjacent modules. On the other hand, the serial AF is controlled by a high resistor between the modules. In contrast, the insertion of a low resistor between the adjacent modules simulates the DF.

Moreover, the PSC is induced by adjusting the irradiance of the PV modules to low levels. The investigated faults and the combination of conditions employed for data collection are concisely outlined in Table 3. Finally, eight features are derived from I to V curve characteristics, FF, G, and T, under normal and faulty operation scenarios to construct the dataset.

### 3.2. Data pre-processing

The next stage is to pre-process the datasets to optimize data processing and improve the outcomes of the CNN model. In this study, missing values in records are handled, categorical labels are converted into numerical representation, and features are normalized to a standard scale.

The normalization of the datasets is achieved by scaling each attribute to a range from zero to one. Data normalization is achieved using a Min-Max normalization technique, as seen in Eq. (4) [6].

$$\text{Normalization}(Min - Max) = \frac{X_i|j - \text{Min}(j)}{\text{Max}(j) - \text{Min}(j)} \quad (4)$$

where  $X_i|j$  is the  $i^{\text{th}}$  data point at feature  $j$ . Meanwhile,  $\text{Min}(j)$  and  $\text{Max}(j)$  denote the minimum and maximum values of feature  $j$ , respectively.

### 3.3. Feature selection

Feature selection approaches aim to identify and select a particular group of relevant features within a dataset to improve the design of learning algorithm models [31]. The redundant and irrelevant features cause extended training time and diminished model accuracy [32]. Consequently, employing feature selection methods is crucial for enhancing the performance of classification problems, as it removes irrelevant features from the dataset. In this study, we have used the permutation importance (PIMP) as a feature selection method. The PIMP method is used to evaluate the importance of features by utilizing the random forests model. This method entails randomly permuting the values of a feature while leaving other features constant to simulate the impact of eliminating that feature from the dataset. The model's performance metric is re-evaluated, where the significance of a feature is determined by the extent of the decrease in performance [33]. This procedure is repeated for all features, enabling the determination of their ranking based on their influence on the model's predictions. An advantage of the random forest algorithm is its ability to easily quantify the relative importance of different features in determining the output.

**Table 3**

The proposed variable operating conditions to collect the datasets.

No	Status of PV array	Status classification	Irradiation	Temperature	Shading	Fault impedance
1	Normal operation	Normal	100:25:1200	0:5:55° c	—	—
2	Shaded three modules	PSC	100:25:1200	0:5:55° c	20 %:20 %:80 %	—
3	Short circuit between two modules	LLF	100:25:1200	0:5:55° c	—	0:15:30 Ω
4	Lose connection between adjacent modules	AF	100:25:1200	0:5:55° c	—	25:25:100 Ω
5	Disconnecting between two modules	OCF	100:25:1200	0:5:55° c	—	1000 Ω
6	Increase in series resistance inside a module	DF	100:25:1200	0:5:55° c	—	4:2:8 Ω

### 3.4. FDD

In this section, the FDD process is divided into two stages: the first involves constructing and developing the classification model, while the second focuses on learning the classification model to classify faults accurately.

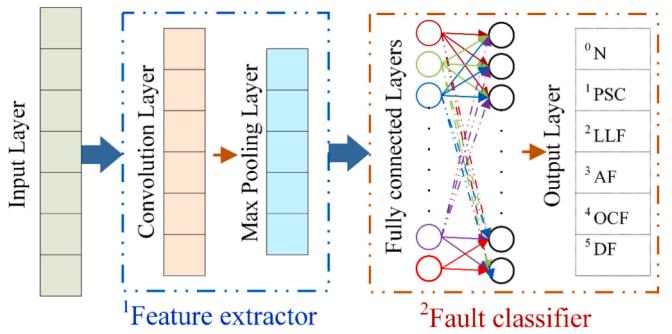
#### 3.4.1. Classification model

In AI model-based method, CNN is widely used for object detection due to its superior performance and data classification abilities [28,34,35]. The 2D-CNN has been used for image processing, while the 1D-CNN has been used for time-series data processing [36–38]. In this study, we proposed a 1D-CNN model to detect and diagnose the faults. The configuration of the proposed model is illustrated in Fig. 10, which consists of a single convolutional layer, a max pooling layer, two fully connected layers, and an output layer. Generally, the model performs two basic tasks: (1) a feature extraction, which is automatically achieved by the learning process from the input data by the convolution and max pooling layers. (2) a fault classification; in this step, the fully connected layers and the output layer are used to classify the faults based on the learned features.

#### 3.4.2. Learning the classification model

The process of learning the 1D-CNN model to extract the features and classify the faults is illustrated in Fig. 11, and it is explained in the following steps:

- 1 Divided the PVA dataset into training, validation, and testing datasets.
- 2 **Training the model:** in this step, the model learns from the training dataset, adjusting its weight and parameters to minimize the error and maximize accuracy.
- 3 **Validating the model:** After each training cycle (or epoch), the model's performance is evaluated on the validation dataset to assess classification accuracy, and then fine-tune the model's hyperparameters based on its performance on the validation dataset to optimize model performance.
- 4 The training and validating steps are iterated many times to cover all the features in their datasets.



**Fig. 10.** The proposed 1D-CNN model.

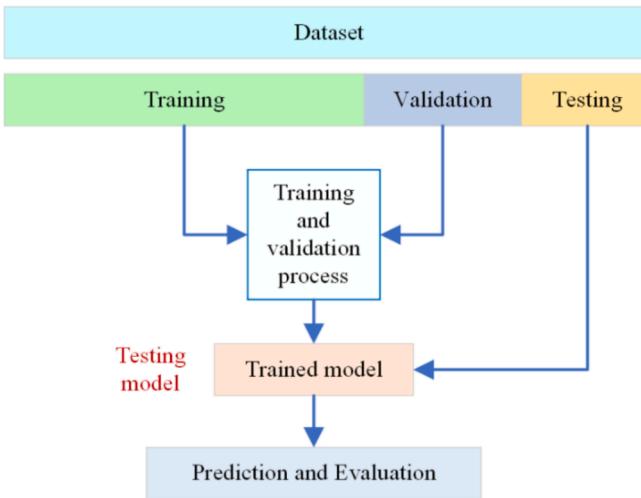


Fig. 11. Flow chart of the training and testing process.

**5 Testing and Validating Model on Unseen Data:** After training and hyperparameter tuning, the trained model is evaluated on the test dataset to confirm its generalization to new, unseen data.

### 3.5. Evaluation metrics

The evaluation metrics in ML/DL are employed to attain an optimized classifier during the training and testing phases of a typical data classification process. The confusion matrix indicates the discriminatory evaluation of the most favorable solution [39,40]. Therefore, we establish the confusion matrix for the proposed classifier to exhibit the correlation between the predicted and actual labels. Then, we use the most widely recognized metrics for evaluating classification performance, such as accuracy, sensitivity or recall, and precision, to assess the proposed classifier based on the established confusion matrix, as shown in the following equations.

- Accuracy is determined by computing the proportion of accurately predicted classes to the overall quantity of assessed samples, as expressed in Eq.5.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

- Sensitivity or recall is employed to determine the proportion of positive instances that are accurately identified, as expressed in Equation (6).

$$\text{Sensitivity(Recall)} = \frac{TP}{TP + FN} \quad (6)$$

- Precision is employed to compute the positive instances accurately forecasted by all predicted instances within a positive class, as expressed in Equation (7).

$$\text{Precision} = \frac{TP}{TP + FP} \quad (7)$$

In the context of the confusion matrix, four critical variables are identified: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). True Positive (TP) refers to the instances where positive cases are correctly identified as positive. True Negative (TN) denotes the cases where negative instances are accurately classified as negative. Conversely, False Positive (FP) represents the instances wherein negative cases are erroneously classified as positive. Lastly, False Negative (FN) captures the scenarios in which positive instances

are incorrectly identified as negative.

## 4. Results and discussion

The proposed models are run on a desktop computer (LENOVO) equipped with a Core™ i5-8400 CPU operating at 2.80 GHz (6CPUs) and 8 GB of RAM, Windows 10 64-bit. The proposed algorithm is implemented in Jupyter Notebook using Python programming language.

### 4.1. Experiment setup and dataset

In this section, a PV system has been simulated to generate datasets for normal and faulty conditions under variable conditions. A PVA model is constructed in MATLAB/Simulink to collect the testbench data from three PVA configurations, including SP-PVA, S-PVA, and TCT-PVA, according to the description in section 3.1. The proposed SP-PVA and TCT-PVA configurations comprise 20 modules distributed across four strings, while the S-PVA configuration is constructed from 10 modules connected in series, as shown in Simulink models in Figs. 12–14. The PVA model has current and voltage sensors to gather the current and voltage. The parameters for the PV module under the standard test condition (STC) are concisely outlined in Table 1. The experiment has different fault configurations, including PS condition on three modules, LLF in two modules, AF between two adjacent modules, OCF on one PV string, and DF between two modules. In this stage, 8640 samples and their respective labels are obtained for six operating scenarios (Normal operation (N), PSC, LLF, AF, OCF, and DF) from three PVA configurations, as given in Table 4.

### 4.2. Evaluation of the proposed FSFD method

This section presents and discusses the experimental results. As described in section 4.1, the datasets were obtained under normal and faulty occurrences utilizing a built data acquisition method. Subsequently, the I-V curve characteristics, FF, G, and T attributes pertinent to the created PVA datasets were retrieved. In total, 8640 data samples have been created and randomly divided into 60 %, 20 %, and 20 % for training, validation, and testing datasets.

In this stage, the created datasets were pre-processed and normalized. The permutation importance method was applied to identify the dataset's optimal features. Figs. 15–17 present the feature importance ranking based on their permutation effect on model accuracy. To select the best features to train the FDD model, four subsets of features were selected based on their ranking in Figs. 15–17, including (1) the top four features, (2) the top five features, (3) the top six features, and (4) the top seven features, for all the proposed datasets. As an example, by using the top four features, the selected features will contain FF, Vmp, T, and Voc for SP-PVA and S-PVA, while they contain FF, Vmp, T, and G for TCT-PVA, as illustrated in Table 6.

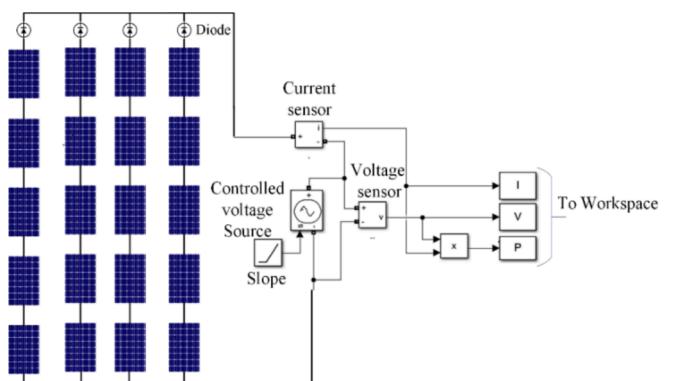


Fig. 12. Simulation model of SP-PVA.

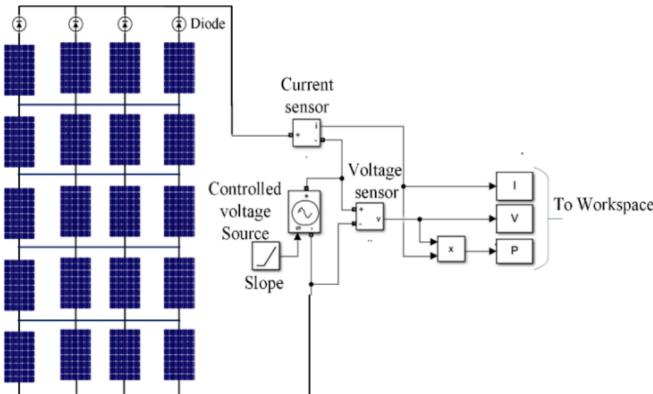


Fig. 13. Simulation model of TCT-PVA.

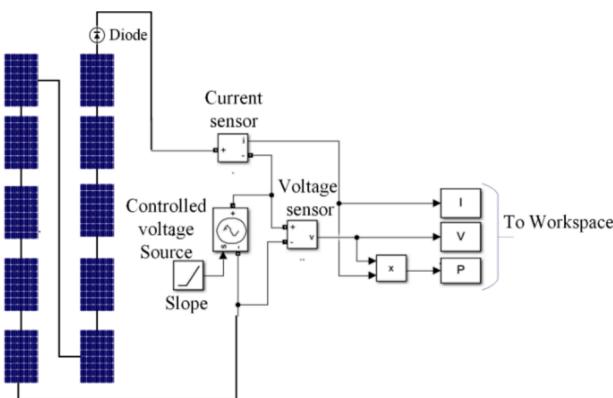


Fig. 14. Simulation model of S-PVA.

**Table 4**  
Dataset description.

Fault Type	Category Label	SP-PVA Samples	TCT-PVA Samples	S-PVA Samples
N	0	540	540	540
PSC	1	2160	2160	2160
LLF	2	1620	1620	1620
AF	3	2160	2160	2160
OCF	4	540	540	540
DF	5	1620	1620	1620
<i>Total of samples</i>		8640	8640	8640

**Table 5**  
Hyper-parameters for the 1D-CNN model.

Layers	No. of kernels	Kernel Size	Stride	Activation	Output
Input layer	–	–	–	–	n
1D-Conv1	64	2	1	Relu	(n-1, 64)
1D-Maxpool	1	2	1	Relu	(n-3, 64)
Flatten	–	–	–	–	64
FC1	1	50	–	Relu	50
FC2	1	25	–	Relu	25
Output layer	1	7	–	Softmax	m

[number of features (n), number of labels (m)].

After creating the subsets of features, the 1D-CNN model was configured to classify multiple classes to detect and classify the faults into six predefined classes, namely N, PSC, LLF, AF, OCF, and DF, which are labeled as 0, 1, 2, 3, 4, 5 class respectively. As mentioned before, we suggested that the CNN algorithm enhance the fault classification. Therefore, choosing appropriate parameters for 1D-CNN will greatly

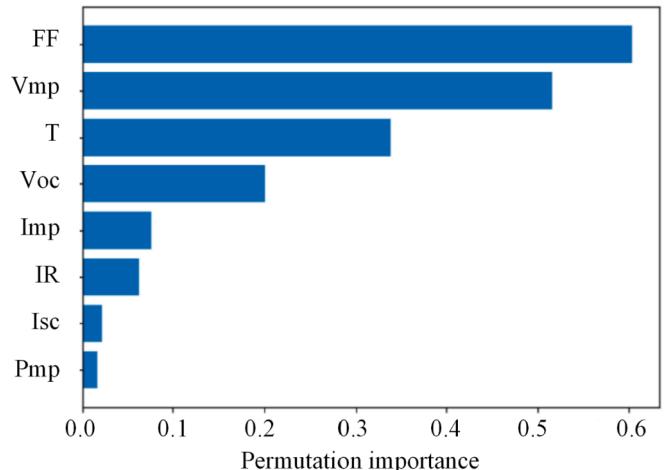


Fig. 15. Feature importance ranking for SP-PVA dataset.

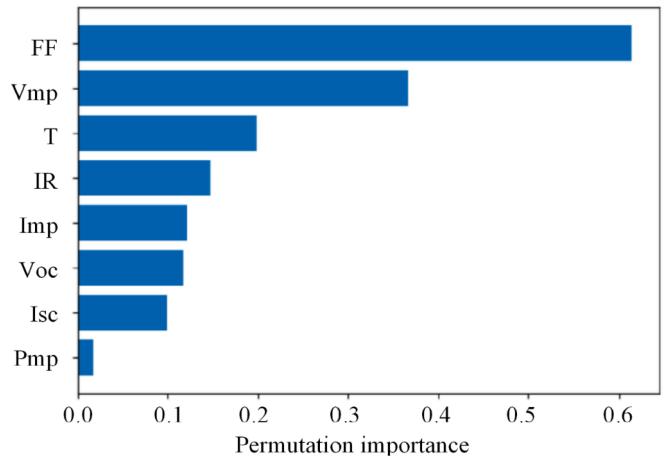


Fig. 16. Feature importance ranking for TCT-PVA dataset.

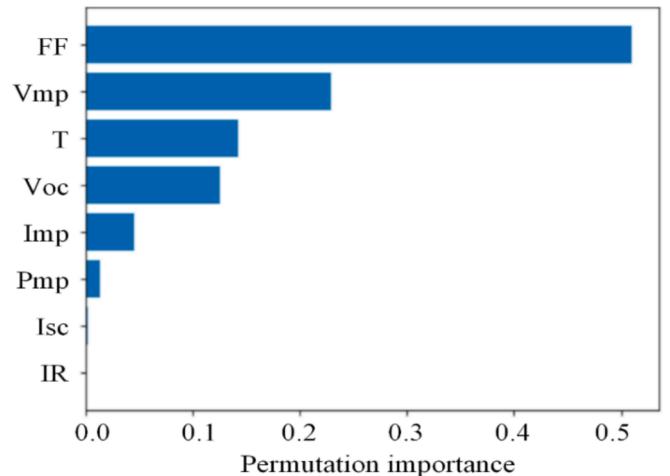


Fig. 17. Feature importance ranking for S-PVA dataset.

impact the classification performance. To do this, we adapted training the 1D-CNN model on the PVA datasets; various combinations of the hyperparameter values were tested, and the most successful combination was chosen. During the training process, the “Adam” optimization technique was used for adaptive moment estimation, and the number of

**Table 6**

Feature selection scenarios based on feature importance rank.

Dataset	Feature ranking							
	1	2	3	4	5	6	7	8
SP-PVA	FF	Vmp	T	Voc	Imp	G	Isc	Pmp
TCT-PVA	FF	Vmp	T	G	Imp	Voc	Isc	Pmp
S-PVA	FF	Vmp	T	Voc	Imp	Pmp	Isc	G
Subset selection	Top four features	Top five features	Top six features	Top seven features				Discarded

training samples (**batch size**) utilized for one iteration was adjusted to **64**. In contrast, the maximum iteration for the training process (epoch) was set to 1000. The hyperparameters and their optimum values are summarized in [Table 5](#).

This study utilizes the hold-out validation approach during the classification stage to measure the training accuracy of the proposed model, which improves reliability and avoids overfitting in the learning process. Consequently, average accuracy and training time are used to assess the training results. To assess the accuracy of the proposed model, The training results, including average accuracy and training time based on the average of four runs of hold-validation, were listed in [Table 7](#).

The results in [Table 7](#) and [Fig. 18](#) indicate that the trained 1D-CNN model with the top five features demonstrates the best performance in training accuracy and time training in all PVA datasets. Specifically, the trained 1D-CNN model on the top five features of the SP-PVA dataset achieved an average accuracy of 99.94 % for training and a training time of 2.23 min. While the trained 1D-CNN model with the top seven features has the same training accuracy, it comes at a higher training time (2.62 min), highlighting a trade-off between accuracy and efficiency. Then, when the TCT-PVA dataset was used to train the model, the trained model with the top five features obtained an average training accuracy of 99.74 % and an average training time of 2.21 min. Finally, with the S-PVA dataset, the trained model with the top five features has achieved the best performance, overcoming the other training cases. It has 100 % training accuracy and 2.20 min for average training time. The analysis indicates that the top five features achieved a balance between model accuracy and efficiency, outperforming the other subsets in detection capability and computational burden. The above findings and investigation indicate that the FSFD methodology can identify the best features, improving accuracy and significantly reducing training duration when used with the 1D-CNN model across different PVA configurations.

Finally, the performance of the trained model was evaluated on the test dataset. The evaluation metrics mentioned in [Section 3.5](#) have been used to evaluate the trained model's performance, including accuracy, sensitivity, and precision. The [Figs. 19–21](#) show the confusion matrixes of the proposed model using the test datasets. The confusion matrix presents the accuracy, sensitivity, and precision. It elucidates precision values for each class in the lowest row. At the same time, the sensitivity metrics are delineated for each class in the far-right column. The cell in the bottom right corner represents the mean overall accuracy across all six labels. The results in [Figs. 19–21](#), indicate that the 1D-CNN classifier can diagnose all the fault labels and the normal label when tested on

three datasets. Firstly, the result of testing the 1D-CNN model on the SP-PVA dataset, as shown in [Fig. 19](#), indicates that Normal, PSC, LLF, OCF, AF, and DF can be identified and classified accurately. Secondly, according to testing the 1D-CNN model on the TCT-PVA dataset, the results in [Fig. 20](#) indicate that PSC, AF, OCF, and DF can be identified and classified accurately. At the same time, Normal and LLF have less sensitivity than other labels.

Moreover, the sensitivity of the Normal category is 98.3 %, and the sensitivity of the LLF category is 98.7 %. In other words, some Normal samples are misjudged as PSC. Similarly, LLF samples are misjudged as the AF category. Lastly, the evaluation of the tested 1D-CNN model on the S-PVA dataset is presented in [Fig. 21](#). The results indicate that the proposed model offers high performance in detecting and diagnosing faults with very high accuracy.

#### 4.3. Evaluation of the proposed model on noisy data

The existence of noise in data, either from the electrical sensors or measurement instruments, impacts the efficacy of diagnostic models. Studies conducted in [\[8\]](#) and [\[41\]](#) utilized white noise to investigate the effect of noise on the classification performance of their proposed protection schemes. In this stage, white Gaussian noise is added to the dataset (clean data) to generate noisy data under varying signal-to-noise ratios (SNR) of 40 dB, 50 dB, and 60 dB to assess the efficacy of the proposed model.

The results in [Table 8](#) reveal that at SNR is equal to 60 dB (reduced noise), the proposed model achieves high accuracy across all datasets, with results of 99.94 % on S-PVA, 99.64 % on TCT-PVA, and 99.68 % on SP-PVA, demonstrating strong performance. As SNR declines to 40 dB (greater noise), the model maintains outstanding accuracy on S-PVA at 99.00 %, indicating strong noise resistance. While on SP-PVA and TCT-PVA, the model displays considerable robustness, with accuracy reducing slightly to 98.31 % and 98 %, respectively. That is to say, even in the case of high noise, the fault classification performance of the proposed model is still at a good level, demonstrating that the proposed model has a good anti-noise ability.

#### 4.4. Evaluating the proposed model on different scales for PVAs

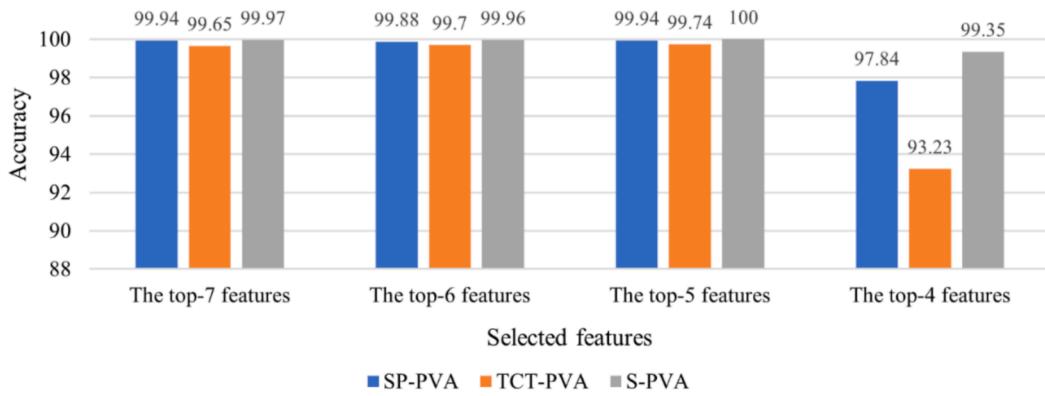
For more validation, the proposed model was tested on different scales for PVAs to assess its sensitivity to variations in network configurations. First, new SP-PVA and S-PVA datasets were created under different operating conditions for LLF and PSC scenarios, including four short-circuited and six shaded modules. Second, the proposed data acquisition method was utilized to record the data from (1) SP-PVA, which was constructed from 30 modules distributed on ten rows and three strings (10x3), and (2) S-PVA was constructed from 13 modules connected in series. Finally, a noise of 50 dB was added to the dataset, and the classification results were recorded in [Tables 9 and 10](#).

The results in [Tables 9 and 10](#) indicate that the 1D-CNN model can classify all the fault and normal classes when tested on the new datasets. Firstly, the result of testing the 1D-CNN model on the new SP-PVA dataset, as shown in [Table 9](#), indicates that Normal, PSC, LLF, OCF, AF, and DF can be identified and classified accurately, the model obtained an overall accuracy of 99.83 %. Secondly, the high overall

**Table 7**

Training results of 1D-CNN model with different feature subsets using three datasets.

Features Subset	SP-PVA		TCT-PVA		S-PVA	
	Training accuracy (%)	Training Time (m)	Training accuracy (%)	Training Time (m)	Training accuracy (%)	Training Time (m)
Top-7 features	99.94	2.62	99.65	2.60	99.97	2.62
Top-6 features	99.88	2.40	99.70	2.42	99.96	2.42
Top-5 features	99.94	2.23	99.74	2.21	100	2.20
Top-4 features	97.84	2.02	93.23	2.03	99.35	2.05



**Fig. 18.** Average training accuracy of the 1D-CNN classifier at different PVA datasets vs. the selected features.

		Sensitivity						
True label	Precision	Normal	PSC	LLF	AF	OCF	DF	
		121 7.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
		0 0.0%	422 24.4%	1 0.1%	0 0.0%	0 0.0%	0 0.0%	99.8% 0.2%
		0 0.0%	0 0.0%	336 19.4%	1 0.1%	0 0.0%	0 0.0%	99.7% 0.3%
		0 0.0%	0 0.0%	0 0.0%	434 25.1%	0 0.0%	0 0.0%	100% 0.0%
		0 0.0%	0 0.0%	0 0.0%	0 0.0%	94 5.4%	0 0.0%	100% 0.0%
	DF	0 0.0%	0 0.0%	0 0.0%	0 0.0%	319 18.5%	100% 0.0%	100% 0.0%
		100% 0.0%	100% 0.0%	99.7% 0.3%	99.8% 0.2%	100% 0.0%	100% 0.0%	99.9% 0.1%
		Normal	PSC	LLF	AF	OCF	DF	Predicted label

**Fig. 19.** Fault classification results for 1D-CNN based on the SP-PVA dataset.

accuracy (99.83 %) of testing the 1D-CNN model on the new S-PVA dataset indicates that the proposed model accurately classifies the classes of Normal, PSC, LLF, AF, OCF, and DF, as shown in the sensitivity's column in Table 10. Finally, the model has a very low misclassification rate, as recorded in the last column in Tables 9 and 10. The results demonstrate that the proposed model accurately detects and classifies faults in different PV network configurations.

#### 4.5. Comparison with other related methods

In this section, the proposed method's efficiency in overcoming the thresholding limitations will be discussed first, followed by a comparison against the AI model-based methods to evaluate the efficiency of the proposed method for diagnosing the PV array's faults. First, the analytical model-based methods in [7–10] reliance on manually predefined thresholds to compare the output power of the PV system to predefined thresholds for detecting and diagnosing two types of faults, including the PSC and SCF in the PV system. The algorithm in [8] has used a single threshold to diagnose PSC and SCF, while the methods in [7,9,10] have utilized multiple thresholds for detecting and classifying the PSC and SCFs in PV modules, PV string and bypass diodes. In contrast, the proposed method offers significant advantages over the

		Sensitivity						
True label	Precision	Normal	PSC	LLF	AF	OCF	DF	
		119 6.9%	2 0.1%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	98.3% 1.7%
		0 0.0%	431 24.9%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
		0 0.0%	0 0.0%	315 18.2%	3 0.2%	0 0.0%	1 0.1%	98.7% 1.3%
		0 0.0%	0 0.0%	0 0.0%	423 24.5%	0 0.0%	0 0.0%	100% 0.0%
		0 0.0%	0 0.0%	0 0.0%	0 0.0%	128 7.4%	0 0.0%	100% 0.0%
	DF	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	306 17.7%	100% 0.0%
		100% 0.0%	99.5% 0.5%	100% 0.0%	99.3% 0.7%	100% 0.0%	99.7% 0.3%	99.7% 0.3%
		Normal	PSC	LLF	AF	OCF	DF	Predicted label

**Fig. 20.** Fault classification results for 1D-CNN based on the TCT-PVA dataset.

analytical model-based methods in automated threshold creation and simplified fault classification. These advantages have made our method more accurate and efficient for detecting and diagnosing PSC and SCF, in addition to three other types of faults, including OCF, AF, and DF, that were not considered in [7–10].

Second, to evaluate the effectiveness of the proposed method against other AI model-based methods, a comparison with methods outlined in [12,16,18,25] has been made, as depicted in Table 11. This comparison included evaluating the accuracy, the number of diagnosed faults, the PVA dataset used to train the classifier, and considering the low-irradiation and high-impedance factors in the collected datasets. The purpose of this comparison is not only to demonstrate the superior diagnostic accuracy of the FSFD method but also to showcase its ideal overall performance.

In comparison to other methods in Table 11, it is evident that the proposed FSFD method demonstrates superior performance. Firstly, the proposed method successfully diagnosed five faults, followed by methods [12] and [16], with the diagnoses of four faults, while methods [18] and [25] detected only three faults. Secondly, the proposed method is tested using three different PVA datasets, including SP-PVA, S-PVA, and TCT-PVA, while method [18] has been tested in two datasets, and the methods [12,16,25] have been tested in a single dataset. Thirdly, the

		Sensitivity						
True label	Normal	120 6.9%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.1%	99.2% 0.8%
	PSC	0 0.0%	431 24.9%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	LLF	0 0.0%	0 0.0%	318 18.4%	0 0.0%	0 0.0%	1 0.1%	99.7% 0.3%
	AF	0 0.0%	0 0.0%	0 0.0%	423 24.5%	0 0.0%	0 0.0%	100% 0.0%
	OCF	0 0.0%	0 0.0%	0 0.0%	0 0.0%	128 7.4%	0 0.0%	100% 0.0%
	DF	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	306 17.7%	100% 0.0%
	Precision	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	99.4% 0.6%	99.9% 0.1%
		Predicted label						
		Normal	PSC	LLF	AF	OCF	DF	

**Fig. 21.** Fault classification results for 1D-CNN based on the S-PVA dataset.**Table 8**  
Classification Accuracy under different levels of noise.

SNR	SP-PVA		TCT-PVA		S-PVA	
	Accuracy (%)	Error (%)	Accuracy (%)	Error (%)	Accuracy (%)	Error (%)
–	99.85	0.15	99.73	0.27	99.97	0.03
60 dB	99.68	0.32	99.64	0.36	99.94	0.06
50 dB	99.44	0.56	99.29	0.71	99.69	0.31
40 dB	98.31	1.69	98.00	2.00	99.00	1.00

proposed method is the only one that considers fault characteristics under low irradiation and high impedance to enhance fault diagnosing. Finally, when comparing the accuracy of the proposed method to those in the studies [12,16,18, and 25], we find that the proposed method has

achieved the highest accuracy of 99.85 % for fault diagnosing in the SP-PVA dataset compared to 99.60 % for the methods [12,18]. For the TCT-PVA dataset, the proposed method has achieved an accuracy of 99.73 %, better than 99.6 % in [18]. On the other hand, the proposed method has achieved the best accuracy of 99.97 % on the S-PVA dataset compared to 99.40 % in [25].

As a result of these comparisons, our method performs best, making it an optimal method for detecting and diagnosing faults in PVAs.

## 5. Conclusion

This paper presents an intelligent FDD method based on feature selection and a 1D-CNN model for detecting and diagnosing faults in PV arrays. The proposed methodology consists of five main stages: data acquisition, pre-processing, feature selection, development of the FDD model, and performance evaluation. The study utilizes an SDM to simulate PV modules and generate datasets under normal and faulty conditions. The datasets are pre-processed, normalized, and subjected to a feature selection method using the PIMP technique to determine the relevant features for training the 1D-CNN model.

The experimental findings illustrate the efficacy of the proposed method across three different PV array configurations: SP-PVA, TCT-PVA, and S-PVA. The 1D-CNN model, trained on the most relevant features, achieved high accuracy rates of 99.94 % for SP-PVA, 99.74 % for TCT-PVA, and 100 % for S-PVA, with minimal training time. The model's performance was further validated on noisy data, showing robust fault classification even under high noise levels (40 dB SNR), with accuracy rates of 98.31 % for SP-PVA, 98 % for TCT-PVA, and 99 % for S-PVA. Additionally, the model was tested on larger-scale PV arrays, maintaining high accuracy and demonstrating its scalability and adaptability to different network configurations.

Compared to existing methods, the proposed method offers several advantages over analytical model-based methods, including automated threshold creation, simplified fault classification, and the ability to accurately diagnose multiple fault types (PSC, LLF, AF, OCF, and DF). Moreover, it outperforms the AI-based methods regarding diagnostic accuracy, the number of detected faults, and the consideration of low-irradiation and high-impedance conditions. The findings demonstrate that the proposed method is a reliable and efficient solution for fault detection and diagnosis in PV systems and can handle various fault scenarios and environmental conditions.

In conclusion, this study contributes to the field of PV system

**Table 9**  
Fault classification results for 1D-CNN based on the SP-PVA (10x3) dataset.

Confusion Matrix						Overall Accuracy	Precision (%)	Sensitivity (%)	Misclassification rate (%)	
Class	N	PSC	LLF	AF	OCF	DF				
N	121	0	0	0	0	0	99.83	99.2	100	0.0
PSC	0	431	0	0	0	0		100	100	0.0
LLF	0	0	317	2	0	0		100	99.4	0.6
AF	0	0	0	423	0	0		99.5	100	0.0
OCF	0	0	0	0	128	0		100	100	0.0
DF	1	0	0	0	0	305		100	99.7	0.3

**Table 10**  
Fault classification results for 1D-CNN based on the S-PVA (13) dataset.

Confusion Matrix						Overall Accuracy	Precision (%)	Sensitivity (%)	Misclassification rate (%)	
Class	N	PSC	LLF	AF	OCF	DF				
N	121	0	0	0	0	0	99.83	99.2	100	0.0
PSC	0	431	0	0	0	0		100	100	0.0
LLF	0	0	319	0	0	0		99.7	100	0.0
AF	0	0	1	421	0	1		100	99.5	0.5
OCF	0	0	0	0	128	0		100	100	0.0
DF	1	0	0	0	0	305		99.7	99.7	0.3

**Table 11**

Comparative the proposed method and related works in [12,16,18,25].

Method	Year	Classifier	diagnosed Faults					PVA dataset type			Fault condition consideration		Accuracy (%)
			PSC	LLF	AF	OCF	DF	SP-PVA	TCT-PVA	S-PVA	Low irradiation	High impedance	
[12]	2022	RUSBoost	✓	✓	✗	✓	✓	✓			✗	✗	99.60
[16]	2022	ANN	✓	✓	✗	✓	✓	✓			✓	✗	93.40
[18]	2020	ANN	✓	✓	✗	✓	✗	✓	✓		✗	✗	99.60 / SP-PVA 99.60 / TCT-PVA
[25]	2024	VPM	✓	✓	✗	✗	✓			✓	✗	✗	99.40
Proposed method	2024	1D – CNN	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	99.85 / SP-PVA 99.73 / TCT-PVA 99.97 / S-PVA

maintenance by providing a robust, intelligent fault detection and diagnosis framework that enhances the dependability and efficacy of PV systems.

Although the detection method has successfully improved the accuracy of diagnosing individual faults in PVA configurations, this study has limitations in diagnosing hybrid faults, which appear when several faults occur simultaneously in the PV system. The main challenge in diagnosing is the resemblance of the hybrid faults' I-V curve characteristics to the I-V curve characteristics of some single faults. Consequently, we recommend further research to enhance the suggested methodology for diagnosing hybrid defects.

#### CRediT authorship contribution statement

**Yousif Mahmoud Ali:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Data curation, Conceptualization. **Lei Ding:** Writing – review & editing, Supervision. **Shiyao Qin:** Supervision, Writing – review & editing.

#### 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.

#### Acknowledgments

This work was supported by the National Science Foundation of China (U22B20101).

#### Data availability

Data will be made available on request.

#### References

- [1] Pillai DS, Rajasekar N. A comprehensive review on protection challenges and fault diagnosis in PV systems. *Renew Sustain Energy Rev* 2018;91:18–40.
- [2] Flicker, J., & Johnson, J. (2013, June). Electrical simulations of series and parallel PV arc-faults. In 2013 IEEE 39th Photovoltaic Specialists Conference (PVSC) (pp. 3165–3172). IEEE.
- [3] Zhao Y, De Palma JF, Mosenian J, Lyons R, Lehman B. Line-line fault analysis and protection challenges in solar photovoltaic arrays. *IEEE Trans Ind Electron* 2012;60(9):3784–95.
- [4] Harrou F, Taghezouti B, Sun Y. Improved \$ k \\$ NN-based monitoring schemes for detecting faults in PV systems. *IEEE J Photovoltaics* 2019;9(3):811–21.
- [5] Pillai DS, Natarajan R. A compatibility analysis on NEC, IEC, and UL standards for protection against line-line and line-ground faults in PV arrays. *IEEE J Photovoltaics* 2019;9(3):864–71.
- [6] Eskandari A, Aghaei M, Milimonfared J, Nedaei A. A weighted ensemble learning-based autonomous fault diagnosis method for photovoltaic systems using genetic algorithm. *Int J Electr Power Energy Syst* 2023;144:10859.
- [7] Iqbal MS, Niazi YAK, Khan UA, Lee BW. Real-time fault detection system for large scale grid integrated solar photovoltaic power plants. *Int J Electr Power Energy Syst* 2021;130:106902.
- [8] Bagherzadeh H, Abedini M, Davarpanah M, Azizi S. A novel protection scheme to detect, discriminate, and locate electric faults in photovoltaic arrays using a minimal number of measurements. *Int J Electr Power Energy Syst* 2022;141:108172.
- [9] Abd el-Ghany HA, ELGebaly AE, Taha IB. A new monitoring technique for fault detection and classification in PV systems based on rate of change of voltage-current trajectory. *Int J Electr Power Energy Syst* 2021;133:107248.
- [10] Dhimish M, Holmes V, Dales M. Parallel fault detection algorithm for grid-connected photovoltaic plants. *Renew Energy* 2017;113:94–111.
- [11] Mouleloued Y, Kara K, Chouder A. A Developed Algorithm Inspired from the Classical KNN for Fault Detection and Diagnosis PV Systems. *Journal of Control, Automation and Electrical Systems* 2023;34(5):1013–27.
- [12] Adhya D, Chatterjee S, Chakraborty AK. Diagnosis of PV array faults using RUSBoost. *Journal of Control, Automation and Electrical Systems* 2023;34(1):157–65.
- [13] Basnet B, Chun H, Bang J. An intelligent fault detection model for fault detection in photovoltaic systems. *Journal of Sensors* 2020;2020:1–11.
- [14] Akram MN, Lotifard S. Modeling and health monitoring of DC side of photovoltaic array. *IEEE Trans Sustainable Energy* 2015;6(4):1245–53.
- [15] Lazzaretti AE, Costa CHD, Rodrigues MP, Yamada GD, Lexinowski G, Moritz GL, et al. A monitoring system for online fault detection and classification in photovoltaic plants. *Sensors* 2020;20(17):4688.
- [16] Voutsinas S, Karolidis D, Voyatzis I, Samarakou M. Development of a multi-output feed-forward neural network for fault detection in Photovoltaic Systems. *Energy Rep* 2022;8:33–42.
- [17] Hussain M, Dhimish M, Titarenko S, Mather P. Artificial neural network based photovoltaic fault detection algorithm integrating two bi-directional input parameters. *Renew Energy* 2020;155:1272–92.
- [18] Ul-Haq A, Sindi HF, Gul S, Jalal M. Modeling and fault categorization in thin-film and crystalline PV arrays through multilayer neural network algorithm. *IEEE Access* 2020;8:102235–55.
- [19] Karmacharya IM, Gokaraju R. Fault location in ungrounded photovoltaic system using wavelets and ANN. *IEEE Trans Power Delivery* 2017;33(2):549–59.
- [20] Dhimish M, Holmes V, Mehrdadi B, Dales M. Comparing Mamdani Sugeno fuzzy logic and RBF ANN network for PV fault detection. *Renew Energy* 2018;117:257–74.
- [21] Khelil CKM, Amrouche B, soufiane Benyoucef A, Kara K, Chouder A. New Intelligent Fault Diagnosis (IFD) approach for grid-connected photovoltaic systems. *Energy* 2020;211:118591.
- [22] 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.
- [23] Hong YY, Pula RA. Detection and classification of faults in photovoltaic arrays using a 3D convolutional neural network. *Energy* 2022;246:123391.
- [24] Chen X, Gao W, Hong C, Tu Y. A novel series arc fault detection method for photovoltaic system based on multi-input neural network. *Int J Electr Power Energy Syst* 2022;140:108018.
- [25] Ding K, Chen X, Jiang M, Yang H, Chen X, Zhang J, et al. Feature extraction and fault diagnosis of photovoltaic array based on current-voltage conversion. *Appl Energy* 2024;353:122135.
- [26] Badr MM, Hamad MS, Abdel-Khalik AS, Hamdy RA, Ahmed S, Hamdan E. Fault identification of photovoltaic array based on machine learning classifiers. *IEEE Access* 2021;9:159113–32.
- [27] 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 IV characteristics. *Appl Energy* 2017;204:912–31.
- [28] Mellit A, Kalogirou S. Artificial intelligence and internet of things to improve efficacy of diagnosis and remote sensing of solar photovoltaic systems: Challenges, recommendations and future directions. *Renew Sustain Energy Rev* 2021;143:110889.
- [29] Qu H, Li X. Temperature dependency of the fill factor in PV modules between 6 and 40 °C. *J Mech Sci Technol* 2019;33(4):1981–6.
- [30] Villalva MG, Gazoli JR, Ruppert Filho E. Comprehensive approach to modeling and simulation of photovoltaic arrays. *IEEE Trans Power Electron* 2009;24(5):1198–208.
- [31] Cai J, Luo J, Wang S, Yang S. Feature selection in machine learning: A new perspective. *Neurocomputing* 2018;300:70–9.
- [32] Zhang T, Zhu T, Xiong P, Huo H, Tari Z, Zhou W. Correlated differential privacy: Feature selection in machine learning. *IEEE Trans Ind Inf* 2019;16(3):2115–24.

- [33] Altmann A, Toloçi L, Sander O, Lengauer T. Permutation importance: a corrected feature importance measure. *Bioinformatics* 2010;26(10):1340–7.
- [34] Abiodun OI, Jantan A, Omolara AE, Dada KV, Mohamed NA, Arshad H. State-of-the-art in artificial neural network applications: A survey. *Heliyon* 2018;4(11).
- [35] Alzubaidi L, 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. *Journal of big Data* 2021;8:1–74.
- [36] Liao GP, Gao W, Yang GJ, Guo MF. Hydroelectric generating unit fault diagnosis using 1-D convolutional neural network and gated recurrent unit in small hydro. *IEEE Sens J* 2019;19(20):9352–63.
- [37] Kiranyaz S, Avci O, Abdeljaber O, Ince T, Gabbouj M, Inman DJ. 1D convolutional neural networks and applications: A survey. *Mech Syst Sig Process* 2021;151: 107398.
- [38] Kiranyaz S, Gastli A, Ben-Brahim L, Al-Emadi N, Gabbouj M. Real-time fault detection and identification for MMC using 1-D convolutional neural networks. *IEEE Trans Ind Electron* 2018;66(11):8760–71.
- [39] Rainio O, Teuhola J, Klén R. Evaluation metrics and statistical tests for machine learning. *Sci Rep* 2024;14(1):6086.
- [40] Ahmadzadeh A, Kempton DJ, Martens PC, Angryk RA. Contingency space: a semimetric space for classification evaluation. *IEEE Trans Pattern Anal Mach Intell* 2022;45(2):1501–13.
- [41] Fahim SR, Sarker SK, Muyeen SM, Das SK, Kamwa I. A deep learning based intelligent approach in detection and classification of transmission line faults. *Int J Electr Power Energy Syst* 2021;133:107102.