

Impact of Fault Resistance on the Performance of ANN-Based Faults Classification in the Farm of PV Systems

Anwer Mahmood Fadhil

Omar Sharaf Al-Deen Al-Yozbaky

Follow this and additional works at: <https://ates.alayen.edu.iq/home>



Part of the [Engineering Commons](#)



ORIGINAL STUDY

Impact of Fault Resistance on the Performance of ANN-Based Faults Classification in the Farm of PV Systems

Anwer Mahmood Fadhil, Omar Sharaf Al-Deen Al-Yozbaky^{ID} *

Department of Electrical Engineering, College of Engineering, University of Mosul, Iraq

ABSTRACT

This research provides an analytical study of the effect of fault resistance on the accuracy of fault classification in solar farms, with the aim of evaluating the performance of classification algorithms in realistic environments where the characteristics of electrical signals resulting from faults change. A model of a 290 kW photovoltaic farm, consisting of 100 strings each containing 7 solar panels, was adopted, and the simulation was carried out using MATLAB/Simulink software. The study included different scenarios for electrical faults, such as line-to-ground (LG) and line-to-line (LL) failures, with the failure impedance changing from 0 to 500 ohms by a step of 20 ohms. Basic electrical signals such as currents, voltages, power, temperature, and radiation, were recorded for each fault. The data was pre-processed, then a classification model was built using the KNIME platform with the Artificial Neural Network (ANN) algorithm, and the data was divided into training and testing groups. The evaluation was conducted using multiple performance indicators: accuracy, retrieval, predictive accuracy, F1-Score, Matthews Correlation Coefficient (MCC), Area Under the Curve (AUC), and Logarithmic Loss (Log Loss), as well as confusion matrices analysis. The results showed that the high fault impedance leads to a gradual decrease in the performance of the model, as a result of the decrease in the clarity of the electrical signals associated with the fault.

Keywords: Photovoltaic systems, Fault resistance, Ground faults, Fault diagnosis, Solar farm monitoring

1. Introduction

Promoting the use of renewable energy sources is a sustainable solution that contributes to mitigating global warming [1]. Photovoltaic (PV) systems are one of the most prevalent technologies due to their low cost and growing interest in environmental issues [2]. Photovoltaic systems can be installed in a variety of places, such as homes, buildings, and utilities, thanks to their flexibility and scalability. Several strategies have been studied to facilitate the effective adoption of these systems, including energy forecasting [3, 4], optimization [5, 6], and surveillance technologies [7]. System failures can lead to significant power loss or serious accidents. For example, at

one location in the UK, there was an 18.9% drop in energy production in one year as a result of failures in the photovoltaic system [8]. Some faults may cause fires or lead to serious problems within power grids [9, 10], as the development of fault detection systems in the photovoltaic system is essential to ensure the stable operation of the systems. The photovoltaic system is one of the most widely used renewable energy systems today, due to its ability to convert solar energy directly into electrical energy without the need for moving mechanical parts [2]. This system is based on the phenomenon of photoelectric conversion that occurs inside solar cells made of semiconductor materials, often silicon, where the absorption of photons releases electrons that produce a constant electric

Received 12 July 2025; revised 13 August 2025; accepted 16 August 2025.
Available online 26 August 2025

* Corresponding author.

E-mail addresses: anwar.23enp30@student.uomosul.edu.iq (A. M. Fadhil), o.yehya@uomosul.edu.iq (O. S. A. D. Al-Yozbaky).

<https://doi.org/10.70645/3078-3437.1042>

3078-3437/© 2025 Al-Ayen Iraqi University. This is an open-access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).

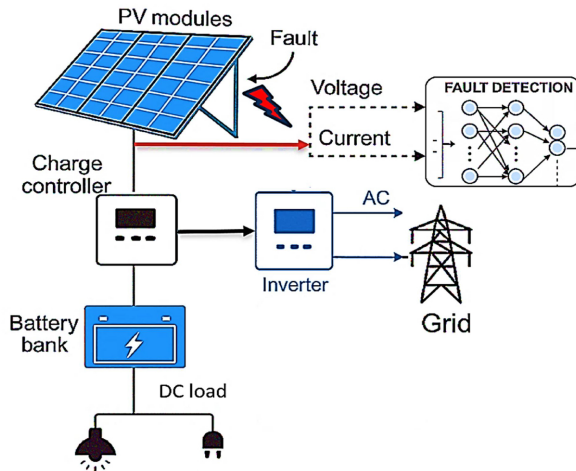


Fig. 1. Grid-connected photovoltaic system architecture with intelligent fault detection mechanism.

current. In its basic form, a photovoltaic system consists of several main components, as in Fig. 1, starting with photovoltaic modules, which are usually installed in groups to form photovoltaic arrays (PV arrays), and then transfer the energy produced to the charge controller, which manages the charging process of batteries and protects them from unsafe operating conditions. In autonomous systems, storage batteries are used to conserve energy for use during periods when the sun is down [11].

The DC power supply from the panels or batteries is then converted into alternating current using an inverter to be compatible with the requirements of electrical loads operating on alternating current (AC). The system is usually equipped with a range of protection and monitoring devices, such as circuit breakers, breakers, grounding systems, and voltage and current measuring devices, to ensure safe and stable operation. Photovoltaic systems can be classified into three main types depending on the way they are connected to the electrical grid [12, 13].

- Off-Grid Systems,
- Grid-Connected Systems
- Hybrid systems that combine autonomy and networking.

A photovoltaic system may experience a variety of failures, including those occurring in the photovoltaic array, power conditioning unit and auxiliary components (Balance of system) [14], As shown in Fig. 2.

The photovoltaic matrix is the only source of fault current, since most PV converters provide galvanic isolation between PV arrays and utility grids, so there is no other source of DC current other than the PV array.

Among these breakdowns, ground faults and line faults have the greatest potential to cause a significant fault stream across the fault path [15]. Without proper fault detection or protection, it can cause serious problems with the PV array, such as DC brackets and even fire hazards [16]. Additionally, sequential or parallel parentheses may occur along these four categories of fault. Especially for string brackets, because they behave similarly to insert variable resistance, which may be difficult to identify or extinguish [17].

Over the past decade, AI techniques have proven effective in modeling, monitoring, simulation, and fault prediction in photovoltaic systems [18]. Several studies have examined the use of (ANN) in the development of custom algorithms for diagnosing faults of these systems [19]. Fore, [20] relied on (ANN) technology in the diagnosis of photovoltaic faults, five different types of faults were analyzed, among which were voltage drops, abnormal energy production, and high temperatures in some components. In another study, [21] used the same technique to locate faults within a single PV string.

Fault impedance is one of the factors that significantly affects the behavior of electrical signals resulting from failures in solar energy [15]. Different

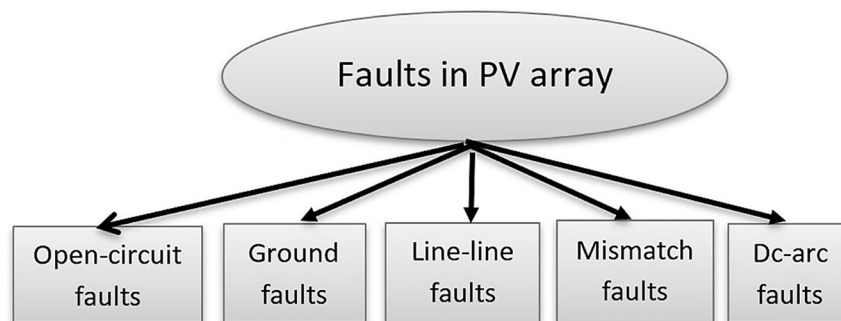


Fig. 2. Common faults in photovoltaic arrays (on the DC side).

impedance values change the range of deviation in current and voltages, which can make it difficult for algorithms to distinguish faults, especially in cases with high impedance [22].

The literature review revealed that many researchers have addressed the issue of fault resistance in photovoltaic (PV) systems from different perspectives. Some focused on fault analysis and examined the impact of fault resistance on electrical parameters, particularly the string current of the affected unit. Others incorporated fault resistance into fault simulations and data collection processes. Additionally, certain studies employed machine learning algorithms to estimate the value of fault resistance, while others aimed to detect faults under high resistance conditions. In contrast, some studies completely ignored this factor, assuming that faults occur as a perfect short circuit with zero resistance—a scenario that does not accurately represent real operating conditions [23–25]. The key aspect that previous works have not addressed—and which forms the core of this research—is the investigation of how fault resistance influences the performance of fault classification algorithms, representing the research gap this study aims to fill.

In this paper, the effect of fault resistance on the classification accuracy of faults in solar farms is investigated to evaluate the robustness of classification algorithms under realistic operating conditions. A simulation model of a 290-kW photovoltaic farm was developed using MATLAB/Simulink, consisting of 100 strings, each containing 7 solar panels. The study focuses on two common types of faults in photovoltaic systems: (LL) and (LG) faults, applied to one or two modules within a string. Fault resistance was varied from 0 to 500 ohms in steps of 20 ohms to simulate different severities of fault conditions. Simulated data were collected for each case, including electrical parameters such as current, voltage, power, along with environmental conditions like temperature and solar irradiance. A total of 3000 samples were generated, with 600 samples for each case and fault type and 600 for normal operation. Random values of temperature (0–50 °C) and irradiance (100–1000 W/m²) were used to reflect real-world variability. After pre-processing, the KNIME platform was used to build a fault classification model using an (ANN). The primary objective of this study is to explore how varying fault impedance influences the performance of intelligent classification systems in practical photovoltaic applications.

The Fig. 1 illustrates an integrated flowchart of a photovoltaic (PV) farm monitoring system with fault detection and classification mechanisms. It combines the MATLAB/Simulink environment for simulating

the PV farm and the KNIME environment for data processing and executing diagnostic algorithms. The system starts from the PV array, which receives solar irradiance (Irr) and temperature (T) as environmental inputs and produces current (i) and voltage (v) as electrical signals. These signals pass through a DC/DC converter equipped with a Maximum Power Point Tracking (MPPT) algorithm and the first control unit, then through an AC/DC converter and the second control unit before connecting to the electrical grid. In parallel, solar irradiance, temperature, current, and voltage are measured using sensors linked to a data acquisition and storage unit. This data is transmitted to the transmission stage and then to the preprocessing unit, where data cleaning and transformation are performed to extract distinctive fault features. Subsequently, the data passes to the fault detection unit, which determines whether a fault exists in the system, followed by the fault classification unit that identifies the fault type and location. Finally, the results, including the system's diagnostic status, are displayed in the visualization unit.

2. Method and tools

Photovoltaic systems sometimes suffer from a range of electrical failures resulting from various internal design problems [26, 27]. This work takes the effect of failure resistance on the accuracy of the algorithms for classifying three different failures, as well as one sound operational state without failures. Fig. 3 shows the types of faults diagnosed in this research with the aim of evaluating the performance of algorithms based on different criteria and comparing their performance to different values of fault impedance. Here

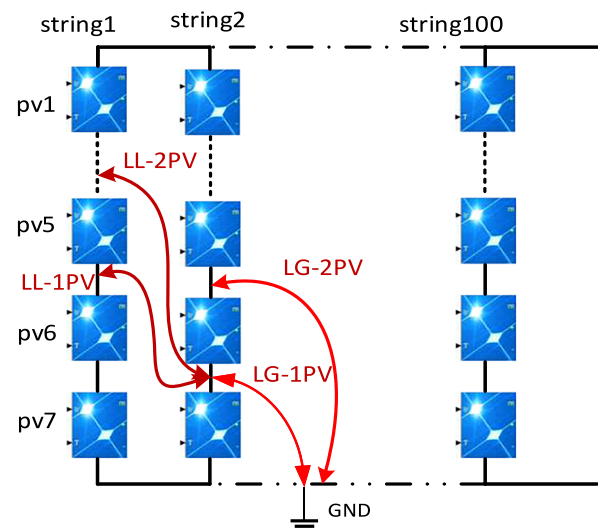


Fig. 3. Describes fault cases LG and LL on an array solar system.

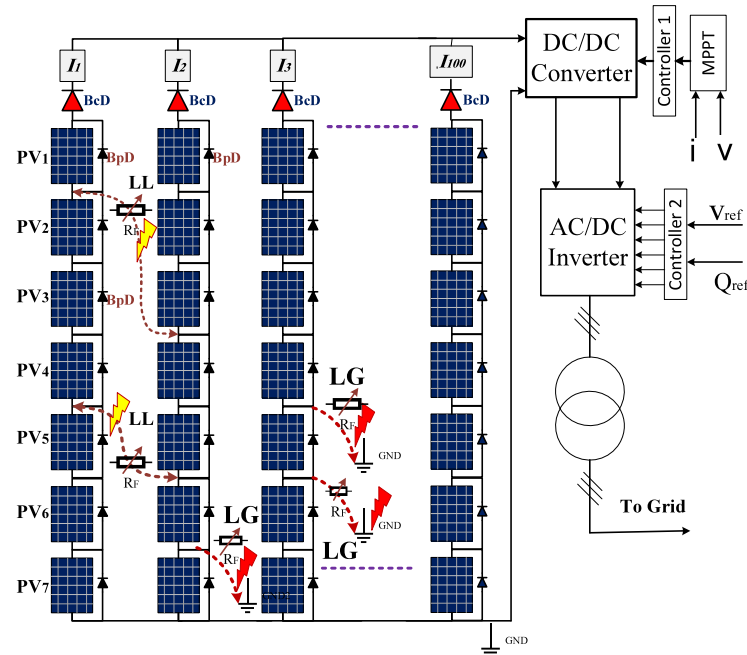


Fig. 4. Describes a 290-kW grid-connected PV farm simulated with Simulink/MATLAB.

is a detailed explanation of the types of faults that occur in the photoelectric arrays.

In this paper, the effect of fault resistance on the performance of the fault classification algorithm in solar systems was studied through an accurate simulation model using the Simulink environment in MATLAB software. The system was represented by simulating a 290 kW solar farm, as in Fig. 4. Fig. 4 represents the simulated model of the solar photovoltaic system, which was developed using the Simulink environment in MATLAB software. The system's connection to the power grid is shown by a three-level inverter based on an IGBT bridge controlled using pulse width modulation (PWM) technology. The inverter controller is designed to apply the maximum power point tracking (MPPT) algorithm using the turbulence and monitoring method. The photovoltaic system was connected to the grid via a three-phase power transformer in a ratio of 0.25/250 kV. The network model includes two short-distance transmission lines: the first line is 14 km long and connected to a 120 kV equivalent network via

a power transformer, while the second line is an 8 km feeder destined for a fixed load. The PV system farm has 100 strings connected in parallel, and each string contains 7 connected modules in series. Each module consists of 128 solar cells, producing a maximum power of 414.801 watts at 72.902 V and 5.6901 amperes, while the open circuit voltage is 85.32 volts, and the short current is 6.091 amperes.

The scenarios are designed to cover three main types of failures: (LL) short, (LG) short, and the following categories. For each fault type, 6 different scenarios were created with gradient fault resistance values: 0, 20, 40, 60, 80, 100, 200, 300, 400 and 500 ohms. Furthermore, faults were applied in two different cases, either across a single solar panel or across two panels described earlier in Fig. 2. In each scenario, 600 sample data were collected, with solar radiation values randomly changing in the range from 100 to 1000 W/m² and temperature within the range from 0 to 60 °C to simulate changing environmental conditions affecting system performance. Categories and number of samples can be described in Table 1.

Table 1. Distribution of 3000 sample training and forecast datasets.

Categories	Count of Samples	Relative Frequency (Samples)	Categories Description
NORMAL	600	%20	No-fault (Normal condition)
LL-1PV	600	%20	(LL) -fault across 1 module
LL-2PV	600	%20	(LL) -fault across 2 modules
LG-1PV	600	%20	(LG) -fault across 1 module
LG-2PV	600	%20	(LG) -fault across 2 modules

Table 1 describes the types of faults that have been simulated for each of the six values of the failure resistance. For each sample, a set of important electrical and environmental variables were recorded, namely temperature, solar radiation, string currents, total system current, total voltage, and total power produced. These data are called features or attributes that represent the basic inputs for analyzing and classifying faults using the algorithm. The research methodology is based on accurate simulation of the system with the application of multiple scenarios for fault resistance values, allowing us to study the impact of this resistance on the accuracy of the fault classification algorithm under changing environmental conditions. This method provides a comprehensive framework for understanding how the system and monitoring systems respond to various faults and enhances the reliability of maintenance and diagnostic operations in solar farms. The data was divided into two separate groups for training and testing purposes by 80% and 20%, respectively. The training kit includes 2400 samples; each sample contains 15 characteristics (feature) in addition to a target variable representing the category (class), as shown in Table 2.

The Fig. 5 illustrates an integrated flowchart of a photovoltaic (PV) farm monitoring system with fault detection and classification mechanisms. It combines the MATLAB/Simulink environment for simulating the PV farm and the KNIME environment for data processing and executing diagnostic algorithms.

The system starts from the PV array, which receives solar irradiance (Irr) and temperature (T) as environmental inputs and produces current (i) and voltage (v) as electrical signals. These signals pass through a DC/DC converter equipped with a Maximum Power Point Tracking (MPPT) algorithm and the first control

unit, then through an AC/DC converter and the second control unit before connecting to the electrical grid. In parallel, solar irradiance, temperature, current, and voltage are measured using sensors linked to a data acquisition and storage unit. This data is transmitted to the transmission stage and then to the preprocessing unit, where data cleaning and transformation are performed to extract distinctive fault features. Subsequently, the data passes to the fault detection unit, which determines whether a fault exists in the system, followed by the fault classification unit that identifies the fault type and location. Finally, the results, including the system's diagnostic status, are displayed in the visualization unit.

After adopting the MATLAB/Simulink simulation environment to simulate failures in the photovoltaic system and extract data related to different failures, and in order to analyze the impact of fault impedance on the performance of fault classification algorithms, six different cases of fault resistance ranging from 0 to 500 ohms were simulated, with a step of 20 ohms. After the data was obtained, it was prepared and processed to be ready for input into the classification algorithm, which was the (ANN) model. The data set was divided into two main parts: the training set and the test set. To build and train the model, the KNIME Analytics Platform was used, which provides a visual and integrated environment for designing and training machine learning models. The workflow is built into KNIME, as shown in Fig. 6, so that raw data is entered, pre-processing steps are performed, and then the neural network model is trained and its performance evaluated at each fault resistance value.

A set of benchmark performance indicators have been adopted to assess the effectiveness of the proposed model in accurately classifying faults, especially under non-ideal operating conditions such as changing failure resistance. Among these indicators, accuracy, which expresses the percentage of properly classified samples, and the F1 index, which is a balanced measure that combines accuracy and retrieval, making it suitable in cases of category imbalances. Matthew's correlation coefficient was also adopted (MCC) as a comprehensive indicator that reflects the overall performance of the model by taking into account all the elements of the confusion matrix, it is considered one of the strongest measures in this context. In addition, other indicators such as Recall, Precision, and Specificity were used to provide detailed insight into the model's ability to distinguish between different states. The (AUC) curve was also adopted as an indicator that measures the model's ability to differentiate between categories in general, along with log loss, which measures the accuracy of the model's prediction of classification probabilities.

Table 2. Description of direct features taken from the photovoltaic system.

Attributes	Data Type	Descriptions
I1mpp	feature	Maximum power point current of string 1
I2mpp	feature	Maximum power point current of string 2
I3mpp	feature	Maximum power point current of string 3
I4mpp	feature	Maximum power point current of string 4
I5mpp	feature	Maximum power point current of string 5
I6mpp	feature	Maximum power point current of string 6
I7mpp	feature	Maximum power point current of string 7
I8mpp	feature	Maximum power point current of string 8
I9mpp	feature	Maximum power point current of string 9
I10mpp	feature	Maximum power point current of string 10
Vmpp	feature	Maximum power point Total DC voltage
Pmpp	feature	Maximum power point Total DC power
Itotal	feature	Maximum power point Total current
T	feature	Temperature (0 °C to 50 °C)
IR	feature	Radiation (100 W/m ² to 1,000 W/m ²)
Category	target	normal, (LG) fault, (LL) fault

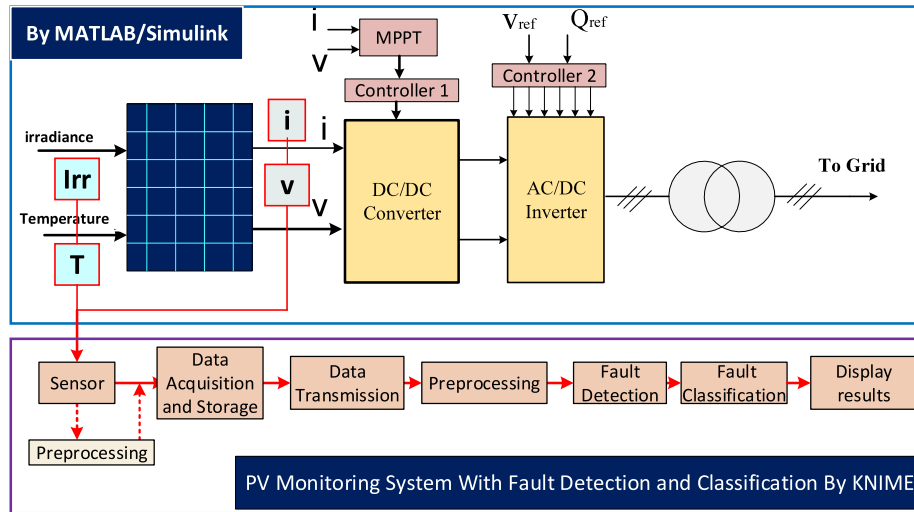


Fig. 5. Flowchart of fault detection and classification system.

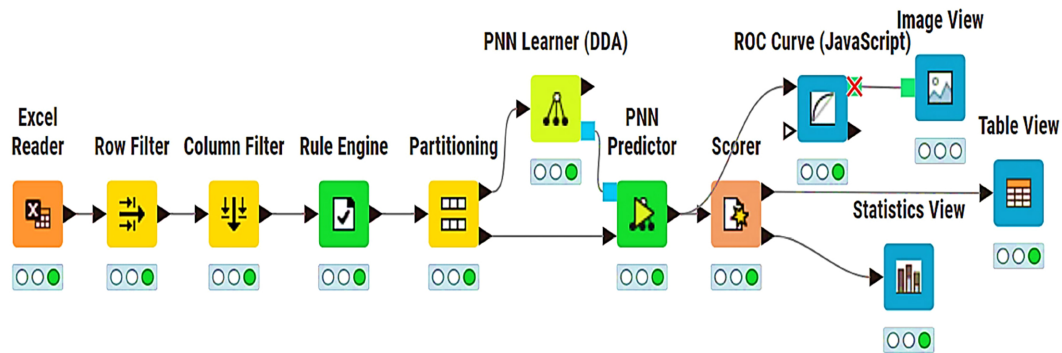


Fig. 6. Fault classification form.

Training Time and Testing Time were also calculated to assess the temporal efficiency of the model, especially in applications that require immediate or near-immediate data processing.

3. Results and discussion

After the completion of the construction of the photovoltaic farm model and the simulation of faults, the focus was on studying the behavior of the electrical outputs of solar strings under the influence of various faults, with a change in fault resistance. The aim of this part is to understand how currents and voltages in fault-affected strings change compared to healthy ones, and how this effect is reflected in the detectability and classification of faults.

Two main fault failures, namely (LG) and (LL) failure, were highlighted, each of which was performed across one unit and two units within the affected string. The fault resistance has been gradually changed from 0 ohms to 500 ohms in irregular

steps, to simulate a range of real-life scenarios that can occur as a result of changing the nature of the contact points or environmental conditions affecting the severity of the failure.

To achieve an accurate understanding of the impact of these failures, the currents of three solar ranges from the farm were mapped and analyzed on the same axes with the aim of direct comparison “As shown in Fig. 7., part (C), the current deviations are more pronounced at 1000 W/m², especially under LL faults involving two modules.” The string (No. 2) was the string in which the failure occurred, while the string (No. 1) was shared in the event of a LL fault., where the fault involves a wire shared between two strings, and the third string (No. 3) is completely intact and is used as a reference for comparison.

This analysis was repeated at three levels of solar radiation: 200 W/m², 500 W/m², and 1000 W/m², in order to monitor the effect of radiation on the clarity of the signals resulting from the failure, where operating conditions and current levels vary according to the intensity of the radiation.

Table 3. Criteria for evaluating the performance of the fault classification model and their description.

Model	Train (sec)	Test (sec)	AUC	CA	F1	Prec	Recall	MCC	Spec	Log Loss	(Rf) Fault Resistance (ohms)
ANN	3.803	0.091	1.000	0.998	0.996	0.998	0.998	0.997	1.000	0.022	0
ANN	3.803	0.091	1.000	0.995	0.995	0.995	0.995	0.992	0.997	0.028	20
ANN	3.803	0.091	0.999	0.980	0.980	0.981	0.980	0.969	0.987	0.246	40
ANN	3.803	0.091	0.999	0.973	0.973	0.974	0.973	0.959	0.982	0.313	60
ANN	3.803	0.091	0.990	0.952	0.952	0.952	0.952	0.925	0.968	0.262	80
ANN	3.803	0.091	0.985	0.947	0.94	0.948	0.947	0.917	0.964	0.345	100
ANN	3.803	0.091	0.976	0.934	0.925	0.940	0.942	0.931	0.951	0.412	200
ANN	3.803	0.091	0.970	0.921	0.920	0.931	0.935	0.924	0.943	0.476	300
ANN	3.803	0.091	0.961	0.913	0.911	0.922	0.927	0.908	0.932	0.514	400
ANN	3.803	0.091	0.953	0.897	0.890	0.901	0.911	0.889	0.921	0.559	500

The figures illustrate three different levels of solar irradiance to show the effect of fault resistance on the currents of three photovoltaic strings under line-to-ground (LG) and line-to-line (LL) fault conditions, applied across one or two modules in the string. In Figure (A), at a low irradiance level (200 W/m^2), the string currents affected by faults start at very low values when the fault resistance is small, with clear differences between the different fault cases. The fault effect is more severe in the LL case compared to LG, and faults affecting two modules cause a larger current reduction than those affecting only one module. As the fault resistance increases, the currents gradually rise and approach the healthy string current until all values converge at high resistance ($\sim 500 \Omega$). In Figure (B), at a medium irradiance level (500 W/m^2), the general pattern is repeated, but the initial current values are higher due to the increased irradiance. The relative differences between the cases are more distinct in the low-resistance range. Again, these differences gradually diminish as the fault resistance increases, almost disappearing at very high resistance values. In Figure (C), at a high irradiance level (1000 W/m^2), the currents in all cases are significantly higher, and the fault effect is clearly visible at low resistances, where the currents of the faulty strings drop considerably compared to the healthy string. However, as the resistance increases, the values converge strongly in all cases, making it nearly impossible to distinguish between them at this point.

From comparing the three figures, it can be concluded that:

- The fault effect is more pronounced at higher irradiance levels, where the absolute current differences between healthy and faulty strings increase.
- The fault type and the number of affected modules have a direct impact on the magnitude of current reduction, with LL faults across two modules being the most severe.

- This pattern has significant implications in the field of fault diagnosis: in the low- and medium-resistance range, the differences between cases are clear and can be relied upon for accurate fault classification, whereas at very high resistances, distinguishing between them becomes extremely difficult because the signals (currents) become nearly identical, which limits the ability of classification algorithms to perform their task efficiently.

The performance of the proposed model was analyzed using a set of statistical criteria previously identified in the methodology chapter. The values of these criteria were calculated after applying the model to the test data to evaluate its accuracy and effectiveness in classifying faults under the influence of variable failure resistance. The criteria used included: accuracy, F1 index, Matthew's correlation coefficient (MCC), positive accuracy, recall, specificity, area under curve (AUC), in addition to log loss, and training and test times. The resulting values of these indicators are organized in Table 3: "Values of model performance evaluation indicators under variable fault resistance", which shows the differences in performance and highlights the efficiency of the model in dealing with different scenarios. The calculated values show the model's ability to balance accuracy and responsiveness, especially in cases with uneven distribution of categories, demonstrating the model's effectiveness in real-world environments.

The results of the performance evaluation of the (ANN) model showed that the value of failure resistance has a clear impact on the accuracy and quality of classification, as the performance was analyzed across a variety of statistical criteria that provide a comprehensive understanding of the efficiency of the model under different operating conditions. Initially, the model showed excellent performance at low crash resistance values, particularly at 0 ohms, where the rating accuracy was 99.8%, with an F1-score of 0.996, demonstrating a high ability to recognize different fault patterns. The precision and recall indicators

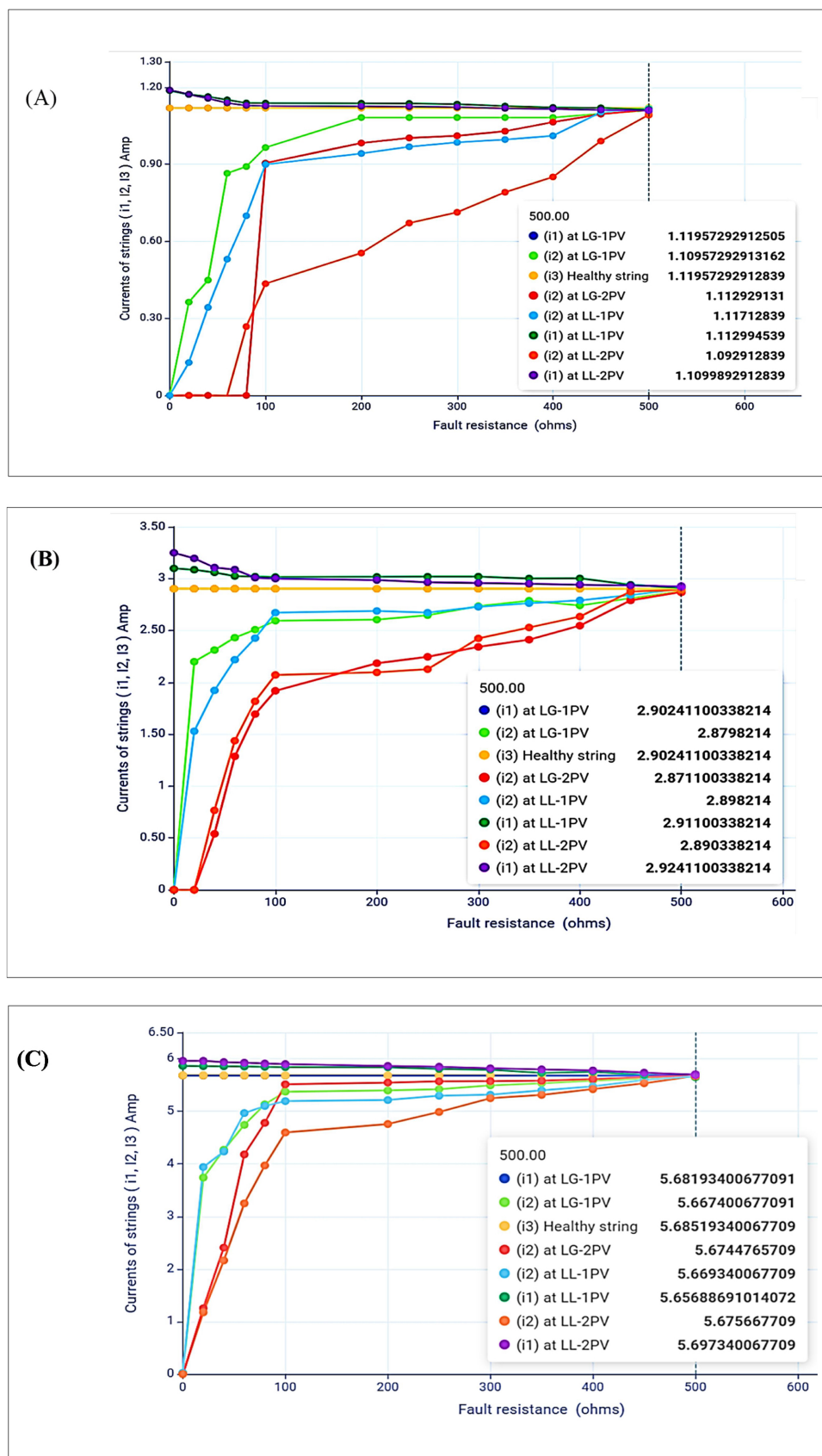


Fig. 7. Variation of currents in 3 PV strings under different fault conditions (LG and LL, across one and two modules) for different levels of solar irradiance, with varying fault resistance: (A) 200 W/m², (B) 500 W/m², (C) 1000 W/m².

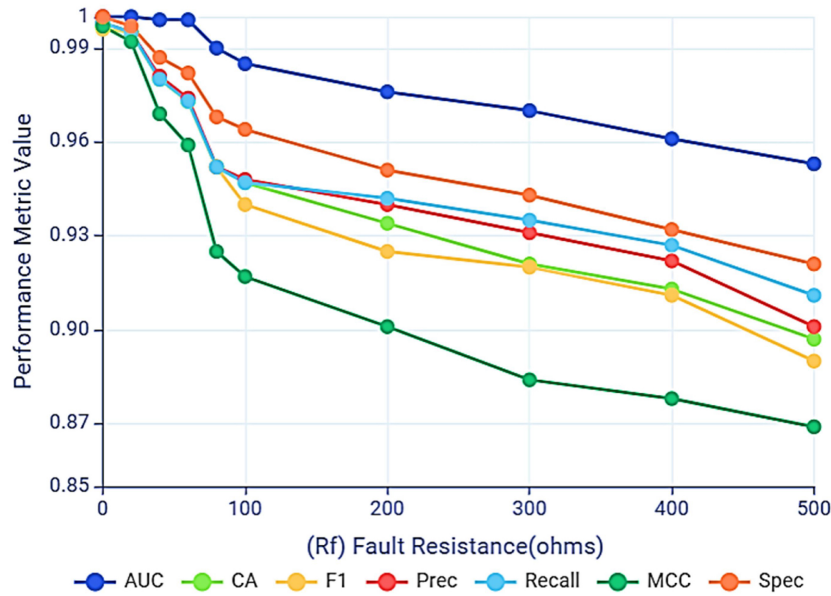


Fig. 8. Graph of the change of the performance parameters of the fault classification algorithm with the difference in the fault impedance value (from 0 to 500 ohms).

were high and proportionate, reflecting a balance in predicting positive categories. This performance distinction can be clearly explained by the electrical signals generated by low-impedance failures.

As the impedance of the fault gradually increased (0–500 ohms), a gradual decline in model performance was observed. Accuracy fell to 0.897%, the F1 score dropped to 0.890 at 500 ohms, and the MCC value fell from 0.997 to 0.889, a strong indicator of the model’s low reliability in complex cases. These results reflect the challenge posed by high fault impedance to the classification model, as the resulting electrical signals become less distinctive.

As shown in Table 3, various criteria such as accuracy, F1-score, MCC, AUC, and log loss illustrate the variation in model performance. For example, specificity remained relatively high, demonstrating the model’s ability to rule out cases without a crash. The value of LogLoss gradually increased, indicating a decline in the model’s confidence in its forecast. Also, the AUC showed a decrease from 0.998 to 0.897, indicating a decline in the model’s ability to distinguish between categories.

Fig. 8, is showed a graphical representation of the change in the performance parameters of the classification model with the change of the fault impedance value from 0 to 100 ohms. From this graph, it can be seen that the model’s performance gradually declines as the impedance value increases, which reflects the inverse relationship between the signal clarity generated by the fault and the algorithm’s ability to accurately classify.

Initially, curves show strong performance at low impedance values, where accuracy is high and a good balance between recall and precision is good, resulting in high F1-score values. The Matthews Correlation Coefficient (MCC) also records high values, demonstrating overall classification quality, especially in multi-class cases. Also, the AUC value is high, indicating the model’s ability to distinguish well between categories at this stage.

As the failure resistance increases, these indicators begin to deteriorate. Accuracy gradually decreases, the F1 score decreases, and the algorithm begins to lose its ability to distinguish subtle patterns in data. The value of the AUC also gradually decreases, reflecting poor model performance in segregating categories. These results confirm that fault impedance directly affects the quality of the electrical signals input to the model and thus the effectiveness of the classification process. Hence the importance of developing algorithms that are more tolerant of subtle changes in signal characteristics or the inclusion of feature engineering techniques to improve performance at high impedances. To understand the fault classification more deeply, confusion matrices were used. A matrix was generated for each fault impedance value (0–500 ohms), as shown in Fig. 9. In matrix (A), at a fault resistance of 0 ohms, the artificial neural network (ANN) demonstrates near-perfect performance, where all fault cases of type LG-Rf0 and LL-Rf0, as well as the normal condition, are classified with very high accuracy. Only one misclassification was recorded, in which an LG case was classified as

		Predicted			Σ
		LG-Rf0.	LL-Rf0.	normal	
Actual	LG-Rf0.	259	0	1	260
	LL-Rf0.	0	240	0	240
	normal	0	0	120	120
	Σ	259	240	121	620
(A)					

		Predicted			Σ
		LG-Rf20.	LL-Rf20.	normal	
Actual	LG-Rf20.	240	0	0	240
	LL-Rf20.	0	239	1	240
	normal	2	0	118	120
	Σ	242	239	119	600
(B)					

		Predicted			Σ
		LG-Rf40.	LL-Rf40.	normal	
Actual	LG-Rf40.	239	1	0	240
	LL-Rf40.	11	229	0	240
	normal	0	0	120	120
	Σ	250	230	120	600
(C)					

		Predicted			Σ
		LG-Rf60.	LL-Rf60.	normal	
Actual	LG-Rf60.	237	3	0	240
	LL-Rf60.	13	227	0	240
	normal	0	0	120	120
	Σ	250	230	120	600
(D)					

		Predicted			Σ
		LG-Rf80.	LL-Rf80.	normal	
Actual	LG-Rf80.	228	12	0	240
	LL-Rf80.	17	223	0	240
	normal	0	0	120	120
	Σ	245	235	120	600
(E)					

		Predicted			Σ
		LG-Rf100.	LL-Rf100.	normal	
Actual	LG-Rf100.	218	22	0	240
	LL-Rf100.	10	230	0	240
	normal	0	0	120	120
	Σ	228	252	120	600
(F)					

Fig. 9. Confusion matrices resulting from fault classification using (ANN) at different fault impedance values: (A) 0 ohms, (B) 20 ohms, (C) 40 ohms, (D) 60 ohms, (E) 80 ohms, (F) 100 ohms.

normal, reflecting a strong ability to distinguish between patterns in the absence of resistance influence.

In matrix (B), at a fault resistance of 20 ohms, the performance remains close to perfect, with a high accuracy in distinguishing the three cases. Only two normal cases were misclassified as LG, and one LL case was misclassified as normal, indicating the beginning of resistance influence, albeit very slightly.

In matrix (C), at a fault resistance of 40 ohms, misclassification becomes more noticeable, with 11 LL fault samples classified as LG, in addition to two other scattered errors. The normal cases, however, continue

to be classified with perfect accuracy. This confusion between LG and LL reflects the reduction in distinctive signal features as resistance increases.

In matrix (D), at a fault resistance of 60 ohms, the confusion persists with a slight increase in errors, as 13 LL cases are classified as LG, and 3 LG cases are classified as LL. Nevertheless, the algorithm maintains almost perfect accuracy in identifying the normal condition.

In matrix (E), at a fault resistance of 80 ohms, the confusion between the two fault types increases further, with 17 LL cases misclassified as LG, and 12 LG

cases misclassified as LL. This indicates that the differences between fault signals are further diminished, making it more difficult to distinguish between the two.

In matrix (F), at a fault resistance of 100 ohms, the confusion between LG and LL reaches its peak, with 22 LL cases classified as LG and 18 LG cases classified as LL. However, the classification of the normal condition remains 100% correct, indicating that the primary challenge for the algorithm lies in differentiating fault patterns at high resistance, while the normal signals remain clearly distinguishable.

Overall, the algorithm's performance is excellent at low resistances but gradually declines as resistance increases, due to the convergence of current patterns between different fault types. This is reflected in the rising confusion between LG and LL, while the accuracy for classifying the normal condition remains constant.

This deterioration in performance can be explained by the fact that high fault impedance reduces the visibility of the resulting signals, making fault characteristics less distinctive, and therefore difficult for the model to distinguish different categories. Based on these observations, it is recommended to enhance the model with more sophisticated classification algorithms, expand the database to include more diverse failures, or even use advanced feature extraction techniques based on signal conversion or spectral processing.

The results of the present work clearly show that fault resistance has a significant impact on the accuracy of fault detection and classification. Nevertheless, many existing studies overlook this factor, which underscores the importance of addressing it when developing reliable diagnostic models for PV systems. The core contribution of this research lies in filling this gap by systematically analyzing the effect of fault resistance on the performance of fault classification algorithms.

4. Conclusion

The results of this research showed that currents extracted from the farm's PV strings were the most affected indicator of faults compared to voltage, which remained relatively constant due to the large number of strings connected in parallel. Increased fault resistance has been shown to reduce current deviation from its normal state, making the electrical signal generated by the fault less noticeable. This decrease in deviation produces overlapping data between normal and fault states, making it difficult for classification algorithms to accurately distinguish

between patterns. different. When analyzing 3,000 samples spread across four scenarios for two major failures (LG) and (LL) failures, across one module and two each), a clear decline in performance indicators was observed with an increase in fault resistance from 0 to 500 ohms. Accuracy decreased from 99.8% to 89.7%, and F1-Score from 0.996 to 0.890, and the coefficient of MCC from 0.997 to 0.889, Graphs have documented this deterioration and showed the inverse relationship between impedance and classification accuracy. The importance of these results highlights the need to take fault resistance into account when designing smart systems to detect faults in solar farms, especially in realistic environments where operating conditions change. The results also show that using current characteristics as the main source of information may be more effective in diagnosing faults at different impedance levels.

5. Recommendations

- Improving the quality of training data: It is recommended to include multiple failures covering a wide range of impedance values, in order to expand the model's generalization capability.
- Adoption of Feature Engineering techniques: Extracting new features or converting signals to improve the model's ability to recognize fine patterns, especially in cases where impedance is high.
- Use hybrid or advanced models: More advanced models such as deep neural networks or hybrid models that combine deep learning with symbolic or statistical reasoning techniques can be tried to improve performance in complex conditions.

Acknowledgements

The authors would like to thank University of Mosul, College of Engineering, Electrical Department, for the support given during this work.

References

1. A. Razmjoo, L. Gakenia Kaigutha, M. A. Vaziri Rad, M. Marzband, A. Davarpanah, and M. Denai, "A technical analysis investigating energy sustainability utilizing reliable renewable energy sources to reduce math altimgemissions in a high potential area," *Renew Energy*, vol. 164, pp. 46–57, Feb. 2021,
2. M. R. S. Shaikh, "A review paper on electricity generation from solar energy," *Int J Res Appl Sci Eng Technol*, vol. V, no. IX, pp. 1884–1889, Sep. 2017, doi: [10.22214/ijraset.2017.9272](https://doi.org/10.22214/ijraset.2017.9272).
3. K. Park, S. Yoon, and E. Hwang, "Hybrid load forecasting for mixed-use complex based on the characteristic load

- decomposition by pilot signals,” *IEEE Access*, vol. 7, pp. 12297–12306, 2019, doi: [10.1109/ACCESS.2019.2892475](https://doi.org/10.1109/ACCESS.2019.2892475).
4. S. Park, S. Yoon, B. Lee, S. Ko, and E. Hwang, “Probabilistic forecasting based joint detection and imputation of clustered bad data in residential electricity loads,” *Energies (Basel)*, vol. 14, no. 1, p. 165, Dec. 2020, doi: [10.3390/en14010165](https://doi.org/10.3390/en14010165).
 5. S. Yoon and E. Hwang, “Load guided signal-based two-stage charging coordination of plug-in electric vehicles for smart buildings,” *IEEE Access*, vol. 7, pp. 144548–144560, 2019, doi: [10.1109/ACCESS.2019.2945483](https://doi.org/10.1109/ACCESS.2019.2945483).
 6. G. Seo, S. Yoon, M. Kim, C. Mun, and E. Hwang, “Deep reinforcement learning-based smart joint control scheme for on/off pumping systems in wastewater treatment plants,” *IEEE Access*, vol. 9, pp. 95360–95371, 2021, doi: [10.1109/ACCESS.2021.3094466](https://doi.org/10.1109/ACCESS.2021.3094466).
 7. J. Song, Y. Lee, and E. Hwang, “Time–frequency mask estimation based on deep neural network for flexible load disaggregation in buildings,” *IEEE Trans Smart Grid*, vol. 12, no. 4, pp. 3242–3251, Jul. 2021, doi: [10.1109/TSG.2021.3066547](https://doi.org/10.1109/TSG.2021.3066547).
 8. S. K. Firth, K. J. Lomas, and S. J. Rees, “A simple model of PV system performance and its use in fault detection,” *Solar Energy*, vol. 84, no. 4, pp. 624–635, Apr. 2010, doi: [10.1016/j.solener.2009.08.004](https://doi.org/10.1016/j.solener.2009.08.004).
 9. L. Hernández-Callejo, S. Gallardo-Saavedra, and V. Alonso-Gómez, “A review of photovoltaic systems: Design, operation and maintenance,” *Solar Energy*, vol. 188, pp. 426–440, Aug. 2019, doi: [10.1016/j.solener.2019.06.017](https://doi.org/10.1016/j.solener.2019.06.017).
 10. C. Correa-Betanzo, H. Calleja, C. Aguilar, A. R. Lopez-Nunez, and E. Rodriguez, “Photovoltaic-based DC microgrid with partial shading and fault tolerance,” *Journal of Modern Power Systems and Clean Energy*, vol. 7, no. 2, pp. 340–349, Mar. 2019, doi: [10.1007/s40565-018-0477-2](https://doi.org/10.1007/s40565-018-0477-2).
 11. A. Rose, R. Stoner, and I. Pérez-Arriaga, “Prospects for grid-connected solar PV in Kenya: A systems approach,” *Appl Energy*, vol. 161, pp. 583–590, Jan. 2016, doi: [10.1016/j.apenergy.2015.07.052](https://doi.org/10.1016/j.apenergy.2015.07.052).
 12. D. D. Milosavljević, T. M. Pavlović, and D. S. Piršl, “Performance analysis of A grid-connected solar PV plant in Niš, republic of Serbia,” *Renewable and Sustainable Energy Reviews*, vol. 44, pp. 423–435, Apr. 2015, doi: [10.1016/j.rser.2014.12.031](https://doi.org/10.1016/j.rser.2014.12.031).
 13. H. N. A. Al-Kaoaz and O. S. A. Y. Al, “Influence of natural clouds on the performance of solar cell systems in Iraq,” *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 4, pp. 1867–1880, 2023.
 14. S. R. Madeti and S. N. Singh, “A comprehensive study on different types of faults and detection techniques for solar photovoltaic system,” *Solar Energy*, vol. 158, pp. 161–185, 2017.
 15. Y. Zhao, J.-F. De Palma, J. Mosesian, R. Lyons, and B. Lehman, “Line–line fault analysis and protection challenges in solar photovoltaic arrays,” *IEEE transactions on Industrial Electronics*, vol. 60, no. 9, pp. 3784–3795, 2012.
 16. J. Flicker and J. Johnson, “Photovoltaic ground fault detection recommendations for array safety and operation,” *Solar Energy*, vol. 140, pp. 34–50, Dec. 2016, doi: [10.1016/j.solener.2016.10.017](https://doi.org/10.1016/j.solener.2016.10.017).
 17. V. Le, X. Yao, C. Miller, and B.-H. Tsao, “Series DC arc fault detection based on ensemble machine learning,” *IEEE Trans Power Electron*, vol. 35, no. 8, pp. 7826–7839, 2020.
 18. A. Mellit and S. A. Kalogirou, “Artificial intelligence techniques for photovoltaic applications: A review,” *Prog Energy Combust Sci*, vol. 34, no. 5, pp. 574–632, Oct. 2008, doi: [10.1016/j.pecs.2008.01.001](https://doi.org/10.1016/j.pecs.2008.01.001).
 19. Y.-Y. Hong and R. A. Pula, “Methods of photovoltaic fault detection and classification: A review,” *Energy Reports*, vol. 8, pp. 5898–5929, 2022.
 20. Yuchuan Wu, Qinli Lan, and Yaqin Sun, “Application of BP neural network fault diagnosis in solar photovoltaic system,” in *2009 International Conference on Mechatronics and Automation*, IEEE, Aug. 2009, pp. 2581–2585. doi: [10.1109/ICMA.2009.5246742](https://doi.org/10.1109/ICMA.2009.5246742).
 21. Syafaruddin, E. Karatepe, and T. Hiyama, “Controlling of artificial neural network for fault diagnosis of photovoltaic array,” in *2011 16th International Conference on Intelligent System Applications to Power Systems*, IEEE, Sep. 2011, pp. 1–6. doi: [10.1109/ISAP.2011.6082219](https://doi.org/10.1109/ISAP.2011.6082219).
 22. M. K. Alam, F. Khan, J. Johnson, and J. Flicker, “A comprehensive review of catastrophic faults in PV arrays: types, detection, and mitigation techniques,” *IEEE J Photovoltaic*, vol. 5, no. 3, pp. 982–997, 2015.
 23. A. Nedaei, A. Eskandari, J. Milimonfared, and M. Aghaei, “Fault resistance estimation for line–line fault in photovoltaic arrays using regression-based dense neural network,” *Eng Appl Artif Intell*, vol. 133, p. 108067, Jul. 2024, doi: [10.1016/j.engappai.2024.108067](https://doi.org/10.1016/j.engappai.2024.108067).
 24. S. S. M. Ghoneim, A. E. Rashed, and N. I. Elkalashy, “Fault detection algorithms for achieving service continuity in photovoltaic farms,” *Intelligent Automation & Soft Computing*, vol. 29, no. 3, pp. 467–479, 2021, doi: [10.32604/iasc.2021.016681](https://doi.org/10.32604/iasc.2021.016681).
 25. K. S. V. Swarna, A. Vinayagam, M. Belsam Jeba Ananth, P. Venkatesh Kumar, V. Veerasamy, and P. Radhakrishnan, “A KNN based random subspace ensemble classifier for detection and discrimination of high impedance fault in PV integrated power network,” *Measurement*, vol. 187, p. 110333, Jan. 2022, doi: [10.1016/J.MEASUREMENT.2021.110333](https://doi.org/10.1016/J.MEASUREMENT.2021.110333).
 26. E. Lodhi *et al.*, “A novel deep stack-based ensemble learning approach for fault detection and classification in photovoltaic arrays,” *Remote Sens (Basel)*, vol. 15, no. 5, p. 1277, 2023.
 27. D. S. Pillai and R. Natarajan, “A compatibility analysis on NEC, IEC, and UL standards for protection against line–line and line–ground faults in PV arrays,” *IEEE J Photovoltaic*, vol. 9, no. 3, pp. 864–871, May 2019, doi: [10.1109/JPHOTOV.2019.2900706](https://doi.org/10.1109/JPHOTOV.2019.2900706).