STM JOURNALS

Journal of
Power Electronics & Power Systems
https://journals.stmjournals.com/jopeps

ISSN: 2249-863X (Online)
ISSN: 2321-4244 (Print)
Volume 15, Issue 2, 2025
May–August
DOI (Journal): 10.37591/JoPEPS

Research JoPEPS

Fault Detection in Solar PV Systems Integrated with the Power Grid: Evaluating Logistic Regression Through Confusion Matrix Analysis

Abhay Nema[1,*], Sparsh Raj[2]

Abstract

This paper proposes a method for failure detection in grid-integrated solar photovoltaic (PV) systems using logistic regression and real-time sensor data. The approach effectively classifies and identifies seven distinct fault types. The developed model demonstrates a high fault identification accuracy, ranging from 93% to 96.5% across various fault types and operational conditions. By leveraging logistic regression, the system utilizes key independent variables that significantly influence the classification process. Additionally, the paper includes a confusion matrix, which offers a visual representation of the differences and similarities between actual and predicted fault states, helping to evaluate the model's performance. The matrix effectively illustrates the model's accuracy in distinguishing between multiple fault categories. This clear classification performance reflects the strength of the proposed method in enhancing fault detection within photovoltaic (PV) systems. By accurately identifying different types of faults, the approach contributes significantly to improving the reliability and operational efficiency of PV installations. Overall, the method demonstrates strong potential in supporting predictive maintenance strategies and reducing downtime, thereby helping ensure consistent performance and longer system lifespan in real-world PV system applications. This method shows promise for real-world applications, where timely and accurate fault identification is essential to maintain optimal performance in solar power generation systems.

Keywords: Solar photovoltaic, PV, logistic regression analysis, real-time sensor data

INTRODUCTION

Until now, much attention has been paid to solar photovoltaics (PV) as a clean, renewable energy source with grid connectivity [1]. Indeed, these systems provide capabilities that are valuable for the global transition to a low-carbon future, particularly in terms of capturing and converting sunlight into electricity [2]. However, solar PV systems are subject to various defects that reduce their performance and longevity; therefore, care should be taken to ensure reliable and efficient operation.

Defect detection and identification have emerged as important research areas in solar PV systems. Equipment failures, climatic conditions, and grid problems result in faults [3–5]. Therefore, it is vital to involve prompt defect identification to ensure that the system performance is maintained, further damage is prevented, and downtime is avoided (Figure 1).

It has been demonstrated that logistic regression is a useful statistical modeling technique for defect identification. Given several independent factors, its application is likely to provide a fault type probability [6]. This, in turn, if the researchers and

***Author for Correspondence**
Abhay Nema
E-mail: abhaynema242@gmail.com

[1,2]Research Scholar, Department of Electrical & Electronics Engineering, Vidhyapeeth Institute of Science and Technology, Bhopal, Madhya Pradesh, India

Received Date: April 15, 2025
Accepted Date: May 02, 2025
Published Date: August 19, 2025

Citation: Abhay Nema, Sparsh Raj. Fault Detection in Solar PV Systems Integrated with the Power Grid: Evaluating Logistic Regression Through Confusion Matrix Analysis. Journal of Power Electronics & Power Systems. 2025; 15(2): 45–52p.

engineers can judge the accuracy and effectiveness of problem diagnosis in PV solar systems through confusion matrix maps and logistic regression model utility [7].

This study focuses on analyzing the confusion matrix resulting from logistic regression to perform seven different fault diagnoses in grid-connected solar PV systems. It measures true positives, true negatives, false positives, and false negatives, thus allowing the evaluation of the classification accuracy for each fault type. The benefits and drawbacks of using logistic regression can then be understood for future improvements in techniques for fault detection, diagnosis, and maintenance.

The primary objective of this study is to investigate the viability of logistic regression analysis as a technique for accurate and reliable fault detection in solar photovoltaic systems. The methodology analyzes the confusion matrix to reveal how the model performs fault classification, enabling a strong basis for appropriate fault detection and mitigation decisions among stakeholders.

This study, through a detailed assessment of confusion matrices, discussed the merits and demerits of logistic regression in the failure diagnosis of grid-tied solar PV systems. This type of information can be effectively employed to develop more reliable and accurate algorithms for fault detection, thereby optimizing the performance of PV installations by reducing maintenance costs.

RELATED WORK

Solar photovoltaic (PV) systems are being integrated into electrical grids with increasing acceptance as a clean and sustainable energy source. It is simultaneously necessary to develop efficient fault detection algorithms for the reliable operation and maintenance of grid-connected PV systems. Several studies have reviewed the detection of faults in large grid-connected PV power plants.

Iqbal et al. proposed an online defect detection system for large-scale solar PV power plants [8]. They developed an intelligent method for recognizing and classifying different fault types using wavelet transformations and artificial neural networks. The study also demonstrated that the proposed system could detect faults accurately and respond to them by reducing their impact.

Chine et al. presented a novel fault diagnosis approach for solar power plants based on artificial neural networks [9]. The study included training the Artificial Neural Network (ANN) to recognize and categorize defect patterns, thus demonstrating the authenticity of ANN applications in fault detection for PV systems, with some research hopeful results about accurate problem diagnosis and classification.

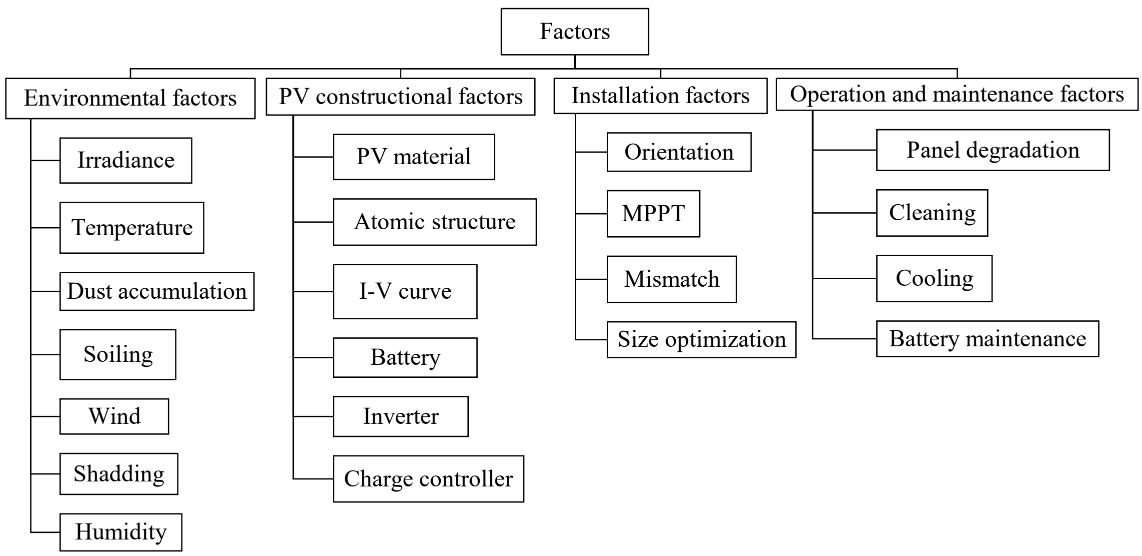

Figure 1. Faults in solar PV systems are due to several factors.

Soffiah et al. studied the faulting of grid-connected photovoltaic (PV) systems using artificial neural networks [10]. The authors established an ANN-based model for the identification and classification of failures based on certain system parameters. Modeling has the potential for real-time fault management in PV systems, and the proposed algorithm was shown to aid in achieving exceedingly high fault detection accuracy. Regarding solar panel fault detection mechanisms, Dhanraj et al. proposed a technique that has been validated as effective [11]. The defect types were identified and classified using machine learning based on current and voltage measurements [12]. The precision and reliability of the fault detection demonstrate the true validation of the proposed method in improving the performance and reliability of PV systems.

The conclusion derived from all the studies is that fault detection has been established as the most important aspect with regard to the dependability of grid-integrated solar photovoltaic systems, with their operation being wholly dependent. ANN-based learning algorithm techniques indicate the promise of modern technologies that are suitable for the identification and classification of a wide variety of fault types. Hence, these detection techniques can provide efficient operation, safety, and long life in terms of locating and rectifying faults, particularly in solar PV systems connected to the grid. Therefore, the domain continues to need further research and development to improve fault detection algorithms and the overall reliability of grid-integrated PV systems.

PROPOSED METHODOLOGY

The maximum power point tracking (MPPT)/incremental perturb and power tracking (IPPT) controller errors in boost converters, open circuits, voltage sags, partial shading, inverters, and current feedback sensors constitute seven distinct fault types.

The following columns are present in every data file:

- *Time:* Time in seconds, average sampling Ts=9.9989 μs.
- *IPV:* PV array current measurement.
- *VPP:* PV array voltage measurement.
- *VDC:* Measurement of DC voltage.
- *IA, IB, IC:* Three-phase current measurements.
- *VA, VB, VC:* Three-phase voltage measurements.
- *IABC:* Magnitude of the current.
- *If:* Current frequency.
- *VABC:* Magnitude of voltage.
- *VF:* Voltage frequency.

Data on faults in a photovoltaic (PV) microgrid system related to the grid were acquired through laboratory tests. Sixteen datasets, each corresponding to a single test, are available in comma-separated values (CSV) and MATLAB (.mat) formats. These datasets include faults in solar clusters, inverters, grid anomalies, feedback sensors, and MPPT regulators, with varying degrees of severity. Information about grid-connected photovoltaic system (GPVS) faults can be used to configure, verify, and evaluate various techniques for fault identification, diagnosis, and characterization in PV system security, protection, and response support.

The MPPT/IPPT operating conditions influence the detection of low-magnitude faults. Severe faults can hinder operation and may cause the system to shut down; therefore, the objective is to detect irregularities before complete system failure occurs. Sample fault data values are provided in Table 1. For formatting purposes, the table is displayed in a transposed form. The target class was calculated using twelve variables that were unrelated to the class. This data sample originated from F0L (no fault) in the low power performance test (LPPT) operational mode. To create a fault condition, a class label is added to each table. In this approach, the emphasis is placed on identifying the fault condition rather than its specific class. Compared with a multiclass classifier, a method for implementing a binary classifier is described.

Table 1. Sample fault data: twelve independent variables, 1 target class.

Time	0.000028178	0.000128	0.000228	0.000328	0.000428
Ipv	1.572327	1.503265	1.492859	1.558136	1.631927
Vpv	101.3489	101.4587	101.5747	101.3123	101.1414
Vdc	144.1406	143.5547	143.5547	143.2617	143.8477
ia	-0.13513	-0.10828	-0.1687	-0.13513	-0.20227
ib	0.490112	0.510254	0.496826	0.510254	0.50354
ic	-0.35499	-0.38855	-0.33484	-0.3617	-0.32142
va	41.74454	46.83151	51.07468	55.84824	60.05524
vb	-149.873	-150.717	-152.019	-152.585	-152.609
vc	109.0646	105.83	102.5431	98.14326	94.26173
Iabc	1	1	1	1	1
If	50	50	50	50	50
Vabc	1	1	1	1	1

The proposed model was examined to determine the most important variables for detecting errors in real-time sensor data. However, the multiclass labelling of the data is not reliable enough to meet a 25% confidence threshold; therefore, only binary classifiers are considered. The dataset from the sensor node follows the framework below for logistic regression. The logistic regression model, based on the linear parameters of the independent variables, yields the following results for R.

Regression Model Summary

Call

Lm (formula: label ~ IPV + VPV + VDC + IA + IB + IC + VA + VB + VC + IABC + IF + VABC + VF, data = train)

Residuals

Min.	1Q	Median	3Q	Max.
-0.175638	-0.005383	0.000086	0.005208	0.097168

Coefficients

```
             Estimate     Std. Error   t value     Pr(>|t|)
(Intercept)  4.415e+00    8.214e-02    53.752      <2e-16 ***
IPV          -5.720e-02   3.014e-04    -189.787    <2e-16 ***
VPV          -3.920e-02   5.703e-06    -6873.722   <2e-16 ***
VDC          -2.528e-03   3.822e-05    -66.146     <2e-16 ***
IA           1.40E-02     9.24E-04     15.15       <2e-16 ***
IB           2.07E-02     8.62E-04     24.038      <2e-16 ***
IC           1.86E-02     9.04E-04     20.563      <2e-16 ***
VA           -9.78E-04    9.83E-06     -99.441     <2e-16 ***
VB           -9.64E-04    9.81E-06     -98.23      <2e-16 ***
VC           -9.60E-04    9.83E-06     -97.633     <2e-16 ***
IABC         3.60E-02     1.25E-03     28.879      <2e-16 ***
If           -1.01E-03    1.22E-04     -8.309      <2e-16 ***
VABC         1.35E-04     5.85E-06     23.023      <2e-16 ***
VF           2.79E-04     1.64E-03     0.17        0.865
```

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ‘ 1
Residual standard error: 0.01099 on 230169 degrees of freedom
Multiple R-squared: 0.9995, Adjusted R-squared: 0.9995

F-statistic: 3.667e+07 on 13 and 230169 DF, p-value: < 2.2e-16

Error in as.data.frame.default(x[[i]], optional = TRUE, stringsAsFactors = stringsAsFactors):
cannot coerce class “summary.lm” to a data.frame

Regression Model Summary

All thirteen parameter values are considerably below 0.05, which represents a sufficiently high confidence level to reject the null hypothesis. This indicates that each of these variables has a significant influence on the computation of fault values.

RESULT AND DISCUSSION

The artificial intelligence model does not rely heavily on parameters such as VDC, IF, IA, IB, and IC; therefore, it is reasonable to omit them.

Figures 2–8 illustrate the confusion matrices of the linear regression analysis for each of the seven defect categories. The matches between the actual and predicted values are shown in green, whereas the differences between the actual and predicted values are shown in red. When evaluated in a clockwise direction, the graph components appear as follows:

1. True Positives
2. False Positives
3. True Negatives
4. False Negatives.

Table 2 and Figure 9 summarize the logistic regression framework created using the complete dataset for both MPPT and IPPT situations.

Table 2. Classification accuracy.

	IPPT	*MPPT*
Fault-1	93	94
Fault-2	95	94
Fault-3	96	96.5
Fault-4	95	94.5
Fault-5	93	93.5
Fault-6	94	95
Fault-7	95	96

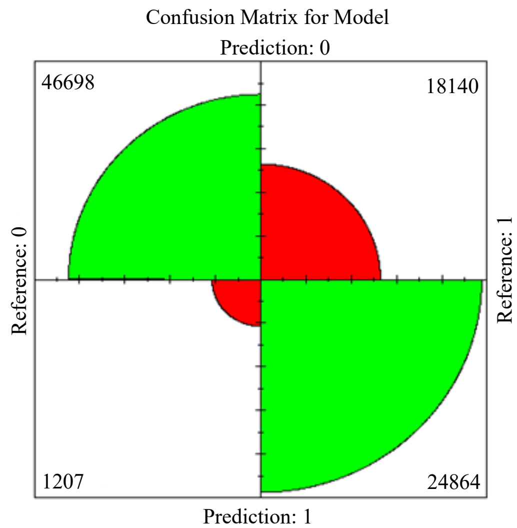

Figure 2. Matrix of confusion: F1.

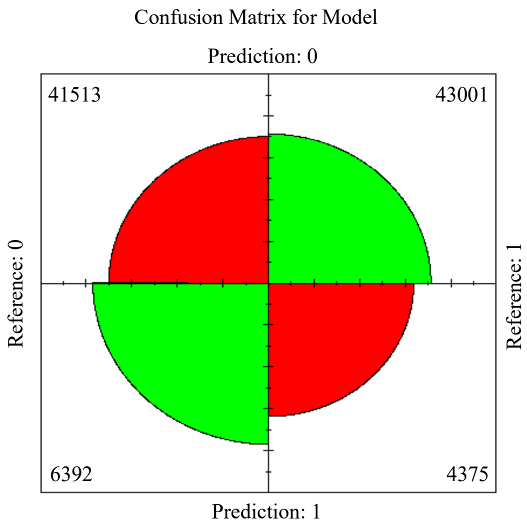

Figure 3. Matrix of confusion: F2.

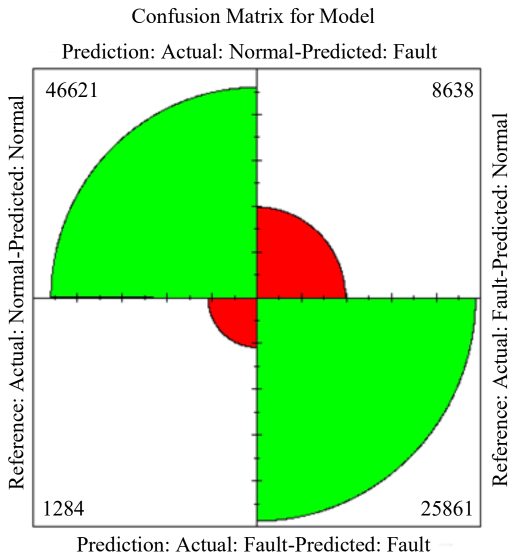

Figure 4. Matrix of confusion: F3.

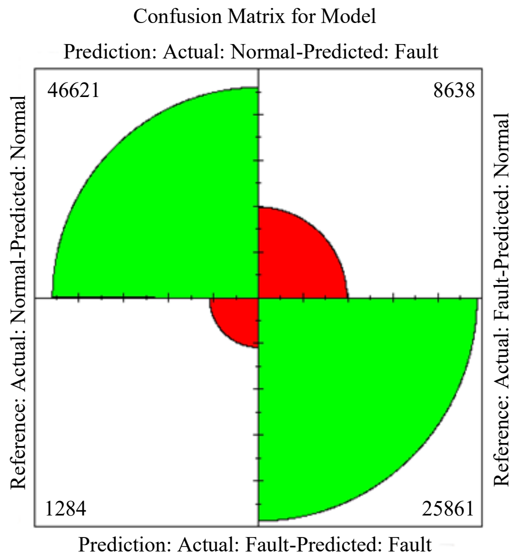

Figure 5. Matrix of confusion: F4.

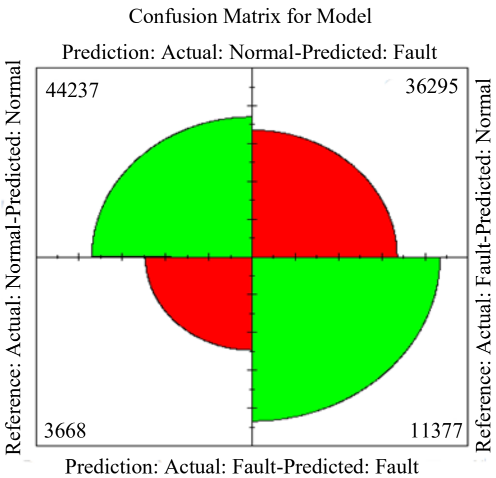

Figure 6. Matrix of confusion: F5.

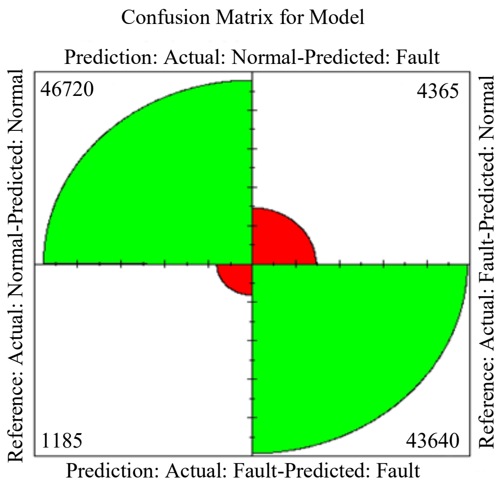

Figure 7. Matrix of confusion: F6.

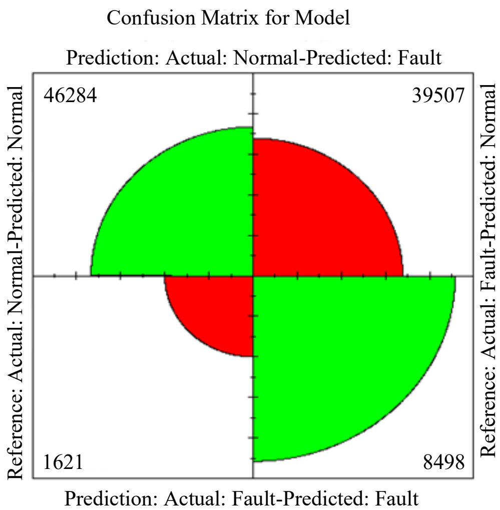

Figure 8. Matrix of confusion: F7.

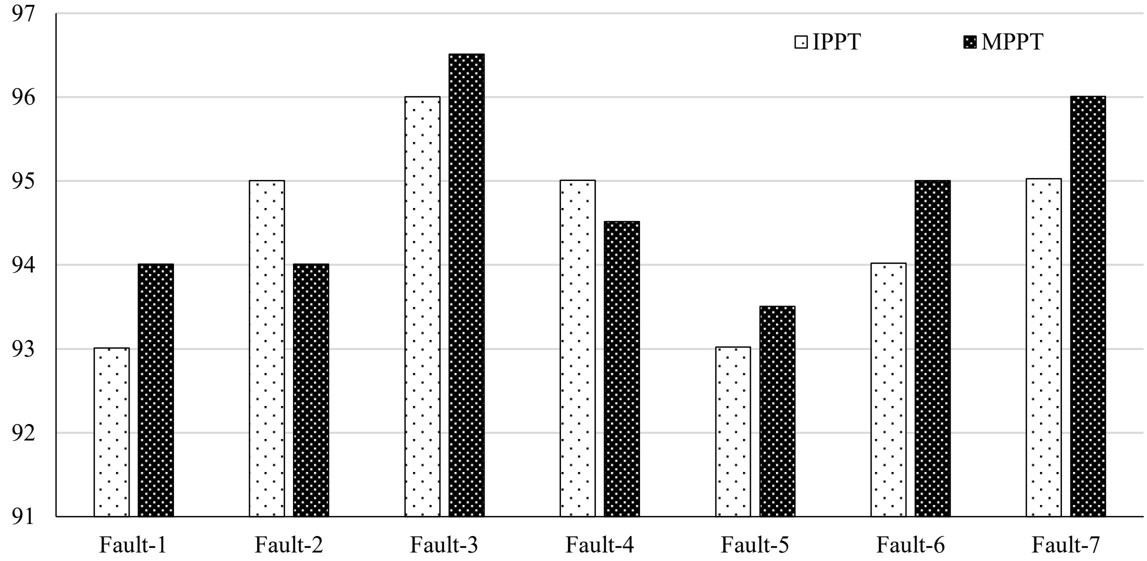

Figure 9. The logistic regression model's accuracy across the seven parameters in relation to IPPT and MPPT values.

CONCLUSION

In conclusion, the proposed logistic regression-based technique shows promise for identifying failures in grid-integrated solar PV systems. This study highlights the importance of selecting appropriate independent variables and demonstrates that some features have a negligible influence on fault identification. The model's accuracy in classifying and identifying the seven fault categories was verified by evaluating its performance using a confusion matrix. The high accuracy rates obtained demonstrate the potential of the proposed methodology for the real-time detection of faults in solar PV systems, thereby enhancing system security and maintenance.

REFERENCES

1. Mellit A, Tina GM, Kalogirou SA. Fault detection and diagnosis methods for photovoltaic systems: A review. Renew Sustain Energy Rev. 2018;91:1–17. doi:10.1016/j.rser.2018.03.062.
2. Jadidi S, Badihi H, Zhang Y. Fault diagnosis in microgrids with integration of solar photovoltaic systems: A review. IFAC-PapOnLine. 2020;53:12091–6. doi:10.1016/j.ifacol.2020.12.763.
3. Triki-Lahiani A, Bennani-Ben Abdelghani AB, Slama-Belkhodja I. Fault detection and monitoring systems for photovoltaic installations: A review. Renew Sustain Energy Rev. 2018;82:2680–92. doi:10.1016/j.rser.2017.09.101.
4. Pillai DS, Rajasekar N. A comprehensive review on protection challenges and fault diagnosis in PV systems. Renew Sustain Energy Rev. 2018;91:18–40. doi:10.1016/j.rser.2018.03.082.
5. Livera A, Theristis M, Makrides G, Georghiou GE. Recent advances in failure diagnosis techniques based on performance data analysis for grid-connected photovoltaic systems. Renew Energy. 2019;133:126–43. doi:10.1016/j.renene.2018.09.101.
6. Nwaigwe KN, Mutabilwa P, Dintwa E. An overview of solar power (PV systems) integration into electricity grids. Mater Sci Energy Technol. 2019;2:629–33. doi:10.1016/j.mset.2019.07.002.
7. Madeti SR, Singh SN. A comprehensive study on different types of faults and detection techniques for solar photovoltaic system. Sol Energy. 2017;158:161–85. doi:10.1016/j.solener.2017.08.069.
8. 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. doi:10.1016/j.ijepes.2021.106902.
9. Chine W, Mellit A, Lughi V, Malek A, Sulligoi G, Pavan AM. A novel fault diagnosis technique for photovoltaic systems based on artificial neural networks. Renew Energy. 2016;90:501–12. doi:10.1016/j.renene.2016.01.036.

10. Sofah K, Manoharan PS, Deepamangai P. Fault detection in grid connected PV system using artificial neural network. 2021 7th International Conference on Electrical Energy Systems (ICEES), Chennai, India. 2021. p. 420–4. doi:10.1109/ICEES51510.2021.9383734.
11. Dhanraj JA, Mostafaeipour A, Velmurugan K, Techato K, Chaurasiya PK, Solomon JM, et al. An effective evaluation on fault detection in solar panels. Energies. 2021;14:7770. doi:10.3390/en14227770.
12. Silvestre S, da Silva MA, Chouder A, Guasch D, Karatepe E. New procedure for fault detection in grid connected PV systems based on the evaluation of current and voltage indicators. Energy Convers Manag. 2014;86:241–9. doi:10.1016/j.enconman.2014.05.008.