Penerbit Akademia Baru

Journal of Advanced Research Design

Journal homepage:
https://akademiabaru.com/submit/index.php/ard
ISSN: 2289-7984

PV Fault Classification: Impact on Accuracy Performance Using Feature Extraction in Random-Forest Cross Validation Algorithm

Muhamad Zahim Sujod[1,*], Siti Nor Azlina Mohd Ghazali[1], Mohd Fadzil Abdul Kadir[2], Ali Qasem Al-Shetwi[3]

[1] Faculty of Electrical and Electronics Engineering Technology, Universiti Malaysia Pahang Al-Sultan Abdullah, 26600 Pekan, Pahang, Malaysia
[2] Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Kampus Besut, 22200 Besut, Terengganu, Malaysia
[3] Electrical Engineering Department, Fahad Bin Sultan University, Tabuk 71454, Saudi Arabia

ARTICLE INFO

Article history:
Received 22 March 2024
Received in revised form 11 November 2024
Accepted 13 December 2024
Available online 31 December 2024

Keywords:
Photovoltaic; fault classification; feature extraction; machine learning; Random-Forest; cross-validation

ABSTRACT

In light of the escalating global concerns regarding energy security and the irregular distribution of daily irradiance affecting photovoltaic (PV) system output, the demand for effective fault detection and diagnosis techniques in PV management systems is on the rise. Machine learning (ML) has emerged as the preferred approach, attracting extensive research attention. As the adoption of solar PV systems continues to surge, the need for robust fault diagnosis and classification techniques becomes paramount to ensure optimal performance, maintenance and scalability across diverse scales of PV arrays in real-world applications. This paper introduces a Solar PV Smart Fault Diagnosis and Classification (SFDC) model that harnesses the Random Forest (RF) algorithm in conjunction with Cross-Validation (CV) and an optimized feature extraction (FE) set. The deployment of CV serves to assess the model's performance and ensure its resilience. Additionally, an optimized FE set is employed to enhance classification accuracy by selecting the most pertinent features for fault classification. Scalable PV array is modelled with the generated power of 10 kW for small-scale, 250 kW for medium-scale and 2 MW for large-scale. In the training and testing of the models, the RF-CV algorithm with set combination of FE was employed to diagnose and classify different types of faults. In this process, each simulated fault that are line-line faults (LLF), open-circuit faults (OCF), ground faults (GF) and partial shading (PS) dataset was divided, with 80% allocated for training and 20% for testing purposes. As the results, all SFDC models (small, medium, large-scale) developed have achieved 100% accuracy for all fault types in training simulations. While in testing the algorithm, it effectively detected multiple fault types, especially OCF and PS, with a perfect score and slightly lower yet high accuracies for GF. These results indicate the robustness and reliability of the SFDC model in diagnosing and classifying faults in PV systems. This research not only holds the potential to advance the field of solar PV for future energy security but also serves as a valuable reference for researchers and policymakers, aiding in the optimization of PV system maintenance and operation.

[*] *Corresponding author.*
E-mail address: zahim@umpsa.edu.my

https://doi.org/10.37934/ard.123.1.6678

1. Introduction

The growing global concern about energy security and the uneven distribution of daily irradiance have exposed the limitations of conventional protection devices (CPD) in effectively detecting and resolving faults in photovoltaic (PV) systems. These undetected faults can persist within the PV array, diminishing system efficiency, reliability and even posing fire hazards [1-3]. These challenges hinder the attainment of optimal performance, maintenance and adaptability in PV systems. There are three primary safety concerns associated with PV systems [4-6]:

i. Shock hazards: These occur when individuals come into contact with uncovered high voltage.
ii. Electrical faults/issues.
iii. Fire risks: Module damage can lead to fires in the PV system.

Among these risks, fire incidents represent the majority of cases and losses to PV systems, structures and operators. The predominant source of fire risk is on the DC side, encompassing string and array cabling, followed by the fault inverter and modules. Incidents of fires on the AC side contribute the least. Poor workmanship is a significant contributing factor to PV system fire accidents, as demonstrated in Figure 1 [7].

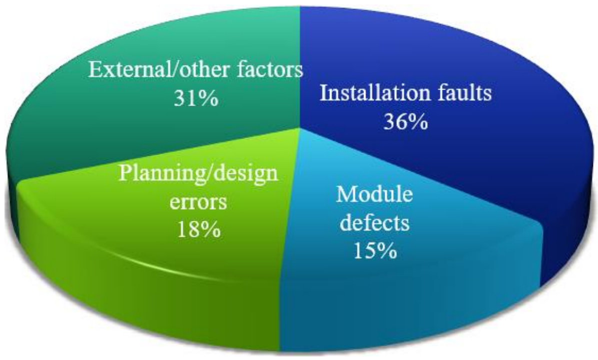

Fig. 1. Factors that contributed to fires

Common faults in PV systems, including partial shading, open circuit faults, ground faults, line-line/short circuit faults, degradation faults and fault modules. Furthermore, common faults in DC and AC systems can be categorized as permanent, intermittent or incipient based on their time characteristics. The attributes of permanent, intermittent and incipient faults are as follows [8-10]:

i. Permanent faults, such as line-line faults, open-circuit faults, ground faults and arc faults, persist until corrected.
ii. Intermittent faults are caused by temporary factors like shading, leaves, bird droppings and environmental stressors such as dust contamination, snow accumulation and high humidity.
iii. Incipient faults result from cell degradation, corrosion and partial interconnection damage. Incipient faults may develop into permanent issues.

All of these faults can endure in PV systems, leading to a decrease in operational efficiency. There is a pressing need for effective fault detection and diagnosis techniques in PV systems to enhance their performance, maintenance and reliability. Therefore, this research aims to contribute to the enhancement of a smart fault diagnosis and classification technique using supervised machine learning algorithm, ultimately improving the accuracy performance and sustainability of PV systems.

2. Literature Review

Machine learning (ML) techniques in fault diagnosis and classification have gained significant attention as powerful approaches to address these challenges. ML techniques have been widely utilized to develop smart fault detection and diagnosis models for PV systems, with the Random Forest (RF) emerging as a popular choice [11-13]. They developed a model in MATLAB/Simulink to detect partial shading (PS), line fault (LF) and open circuit fault (OCF) in unbalanced PV array data. The model has been designed for real-time applications and requires only a simple computational system. The study achieved high accuracy in fault detection and diagnosis. However, it was only validated on a small-scale 2 kW grid-connected PV system. Additionally, RF with supervised algorithm is not been widely studied and implemented in the PV system.

While RF has shown promise in fault diagnosis and classification tasks, its effectiveness and accuracy can be further enhanced by integrating other approaches and optimizing the feature extraction (FE) process. Previous studies by Bacha *et al.,*[14], Akhtar *et al.,* [15], Murtaza *et al.,* [16], Al-Shetwi *et al.,* [17] and Dhibi *et al.,* [18] have found that the most used FE in developing PV fault electrical-related detection and diagnosis methods were P_{max}, Mean, STD, RMSE and variance. These FE methods have been shown to achieve good accuracy for PV system fault detection and diagnosis. However, since the PV output exhibits nonlinear characteristics due to daily changing solar irradiance data, which is influenced by varying meteorological conditions and changes over time, it is recommended to select an ideal time-series FE method.

Therefore, this study endeavours to advance the solar PV fault diagnosis and classification model by leveraging the capabilities of the RF-Cross-validation (CV) with FE. CV technique will be employed to assess the model's performance and robustness and simulation processes using real irradiance data will contribute to enhanced accuracy. The utilization of an optimal, which is the combination of several FE set will further improve fault classification accuracy by selecting the most relevant features. The findings of this research will offer valuable insights to industry practitioners and researchers, aiding in the optimization of maintenance strategies and the overall operation of PV systems. Ultimately, this research will contribute to the wider adoption of solar PV as a clean and sustainable energy source, promoting energy security and environmental sustainability. It will serve as a valuable reference for researchers and policymakers seeking to optimize the maintenance and operation of PV systems.

3. Methodology

3.1 Series and Parallel Configuration of PV Array

Photovoltaic (PV) panels or modules within a PV array are interconnected using parallel, series or a combination of both configurations to achieve the desired power output. This flexibility in arrangement allows for various PV array configurations. The PV array model has the capability to be scalable. The model can be expanded to create PV array configurations with modules arranged in series and parallel arrangement. This scalability feature enables the PV array model to adapt to different sizes and configurations, making it applicable for various real-world scenarios. Detail explanation is written in previous studies by Ghazali *et al.,* [19,20]. Therefore, using the PV array base model, the desired PV array models can be developed and generate simulated data of small, medium and large-scale PV array. In this study, the generated power of PV array for small-scale, medium-scale and large-scale is 10 kW, 250 kW and 2 MW respectively.

The real (measured) PV output data utilized for the simulation processes in this study was sourced from the power management database of the KMSB Solar PV plant located in Pasir Mas, Kelantan, Malaysia. The simulation is using the MATLAB/Simulink program. The simulation process involved utilizing the PV array base model along with input parameters listed in Table 1. For the PV panel input, the specific manufacturer chosen was Panasonic and the model selected was VBMS250AE04, as listed in Table 1. This selection allowed for accurate and reliable simulation of PV power output data, enabling the research to proceed with its objectives effectively.

Table 1
Data for Panasonic VBMS250AE04 PV module

Parameter	Symbol	Value
Maximum Power	P_{mpp}	250 W
Open Circuit Voltage	Voc	37.4 V
Maximum Power Voltage	V_{mp}	30.2 V
Short Circuit Current	Isc	8.86 A
Maximum Power Current	I_{mp}	8.30 A
Diode saturation current	Io	2.75e-10 A
Diode ideality factor	N	1.0136
Shunt resistance	R_{sh}	inf
Series resistance	R_s	0.15 Ω
Solar cell number in series	n	48

3.2 Training and Testing of RF-CV Algorithm Procedure

In the training and testing of Smart Fault Diagnosis and Classification (SFDC) model, the RF-CV algorithm was employed to diagnose and classify different types of faults. In this process, each simulated fault that are line-line faults (LLF), open-circuit faults (OCF), ground faults (GF) and partial shading (PS) dataset was divided, with 80% allocated for training and 20% for testing purposes. The flowchart of this procedure is illustrated in Figure 2.

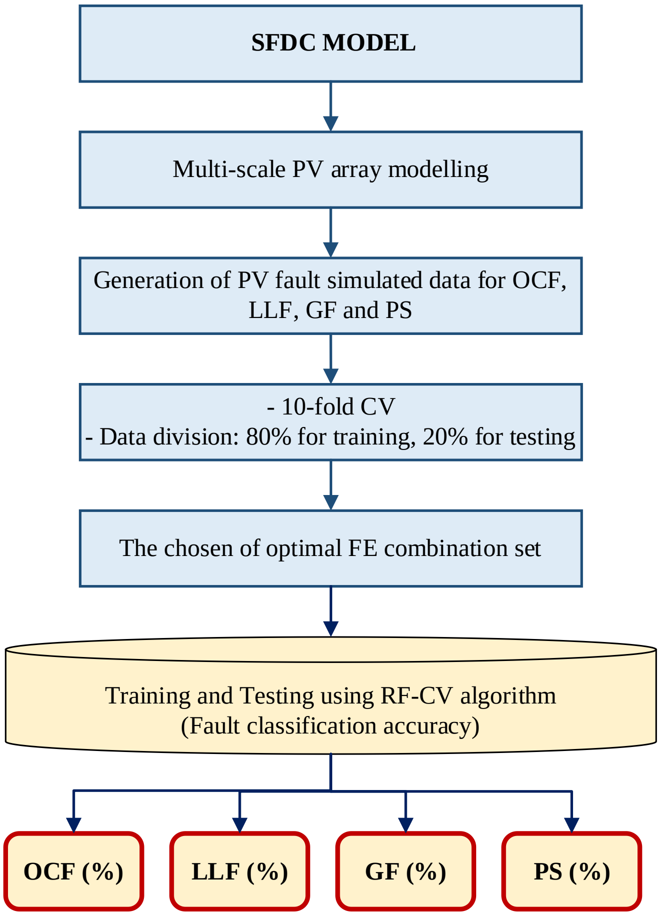

Fig. 2. The flowchart of training and testing SFDC model

Moreover, the RF-CV algorithm was integrated with the 10-fold CV in training and testing process, where during this process, the data was evenly partitioned into ten subsets, with each subset sequentially used for testing while the remaining nine subsets were utilized for training the classifier. Ultimately, the mean accuracy across the ten subsets was recorded. Furthermore, the training and testing process of the RF-CV algorithm was repeated for the medium and large PV array models. This repetition is crucial in this research to develop the feasible SFDC model for multi-scale PV arrays.

3.3 The Chosen of Suitable Feature Extraction (FE)

Feature Extraction (FE) is a critical component in algorithm training and testing process, ensuring that the proposed algorithm functions effectively and produces desirable outcomes [21,22]. In this study, a simulation to examine the effect of individual FE is conducted on the each of PV fault classification accuracy. Subsequently, the performance of combination FE was examined and the FE combination set that produced the best results was chosen to be employed in the training and testing RF-CV algorithm. This research evaluated seven FEs which have proven produced good accuracies in detecting and diagnosing faults in PV systems.

The definition and mathematical expression of the selected FEs in this research are as the following [23-26]:

i. Power maximum (P_{max}) referring to the point on the I-V curve where the generated product of current and voltage is maximum.
ii. Mean refers to the average PV power output.

$$(M) = (\overline{x}) = \frac{\sum_i^n x_i}{n} \tag{1}$$

iii. Standard deviation measures how dispersed the PV power output data is in relation to the mean.

$$(STD) = \sqrt{\frac{\sum_i^n (x_i - \overline{x})^2}{n-1}} \tag{2}$$

iv. Root mean square refers to the square root of the mean square of the PV power output data set.

$$RMS = \sqrt{\frac{1}{n}\sum_n x_i^{\,2}} \tag{3}$$

v. Skewness measures the distortion of the symmetrical distribution of a PV power output data set.

$$Skew = \frac{\sum_n^n (xi - \overline{x})^3}{(n-1) \times STD^3} \tag{4}$$

vi. Waveform length (WL) feature was used and explored in this research's training and testing algorithm process because the PV output data is characterized by time series and has non-linear characteristics caused by varying meteorological influences and changing solar radiation.

$$WL = \sum_i^n |x_i - x_{i-1}| \tag{5}$$

vii. Autoregressive (AR) has been investigated in the previous study that involved time series analysis and proven obtained good results.

$$AR = \sum_i^n \varphi i X_{t-i} + \varepsilon_t \tag{6}$$

Where x_i is the PV power output (PVPO) data and n is the sum of the PVPO and φ_1,…..φ_n are the model parameters and ε_t is the white noise.

4. Results

4.1 Impact of Feature Extraction (FE) on Training and Testing with RF-CV Algorithm (Small-Scale PV Model)

The results and analysis concerning the effect of applying the ideal FE on the training and testing simulations of the RF-CV algorithm have proven efficiently enhanced the precision and accuracy of the SFDC model’s fault diagnosis and classification. The average accuracy of fault classification, obtained from the training and testing simulations of the RF-CV algorithm for the no-fault model of the small-scale PV array, along with the proposed seven features, is presented in Table 2.

Table 2
Average PV fault classification accuracy over seven FE for small-scale PV array model

Feature Extraction	Average Accuracy of PV Fault Classification (%)	
	Training RF-CV algorithm	Testing RF-CV algorithm
P_{max}	100.00	97.00
M	100.00	80.00
STD	98.00	96.67
RMS	90.25	87.00
$Skew$	82.92	78.67
WL	98.00	97.67
AR	100.00	89.33

The classification accuracy is high for all features, with P_{max}, M and AR achieving an accuracy of 100% in the training algorithm process. For the testing simulation of the RF-CV algorithm, on the other hand, the WL feature shows the highest accuracy of 97.67%. Among these features, the *Skew* feature has the lowest accuracy in the testing and training phases. The accuracy list for the proposed combination of FE set, which was trained and tested on the small-scale PV array model is presented in Table 3. The best results produced are highlighted in bold.

Based on the data displayed in Table 3, the proposed combination of features has a high average accuracy for classifying faults in a small-scale PV model using the RF-CV algorithm. The average accuracy ranges from 97.66% to 100.00% for the testing and training data sets. Furthermore, as more features were added to the combination, the accuracy increased, which suggests that the additional features provide useful information for fault classification. All the feature combinations have produced 100% average accuracy for the training RF-CV algorithm. For the testing RF-CV algorithm, on the other hand, the highest average accuracy of 99.00% has been achieved when the combination of (P_{max}+M+STD+RMS+$Skew$), (P_{max}+M+STD+RMS+$Skew$+WL) and (P_{max}+M+STD+RMS+$Skew$+AR) were used in the simulation processes. Thus, it can be concluded that these combination features are particularly important to enhance fault classification in the small-scale PV model.

Table 3
Combination of feature extraction accuracy (Small-scale PV model)

Feature Extraction	Average Accuracy of PV Fault Classification (%)	
	Training RF-CV Algorithm	Testing RF-CV Algorithm
$P_{max} + M + STD$	100.00	97.66
$P_{max} + M + STD + WL$	100.00	97.66
$P_{max} + M + STD + RMS$	100.00	98.00
$P_{max} + M + STD + AR$	100.00	98.33
$P_{max} + M + STD + AR + WL$	100.00	98.33
$P_{max} + M + STD + RMS + Skew$	100.00	99.00
$P_{max} + M + STD + RMS + Skew + WL$	100.00	99.00
$P_{max} + M + STD + RMS + Skew + AR$	100.00	99.00

The accuracy of each fault type can be calculated as the ratio of the diagonal element of that fault type to the total number of instances in that fault type, 75 in this case (number of correctly classified instances in the first fault type)/ (total number of instances in the first fault type). It can be seen that for GF, one output data has been wrongly classified as LLF, while for LLF, two output data have been wrongly classified as GF. For the OCF and PS, the model correctly classified all 75 instances in each fault type, resulting in an accuracy of 100% for both. The accuracy of each fault type resulting from training and testing the RF-CV algorithm is described in detail in Table 4.

Table 4
Fault type classification accuracies (Small-scale PV model)

RF-CV algorithm	Fault type classification accuracy (%)				
	GF	LLF	OCF	PS	Average
Training	100.00	100.00	100.00	100.00	100.00
Testing	98.70	97.33	100.00	100.00	99.00

From the data in Table 4, during training, the RF-CV algorithm achieved a perfect 100% accuracy in classifying all fault types within the training dataset. This means it correctly identified every instance of GF, LLF, OCF and PS in the training data. In the testing work, the RF-CV algorithm also achieved 100% accuracy for OCF and PS. For the GF, the RF-CV algorithm achieved as second higher classification accuracy of 98.7% and follow by the LLF with the classification accuracy of 97.33%. Overall, the RF-CV algorithm performs strongly in fault-type classification for the small-scale PV model. It achieved high accuracy for most fault types, with some minor variation in performance across different fault types in the testing dataset.

4.2 Impact of Feature Extraction (FE) on Training and Testing with RF-CV Algorithm (Medium-Scale PV Model)

Table 5 presents the average accuracy of fault classification for the medium-scale PV array model, obtained from the training and testing simulations of the RF-CV algorithm, employing seven features. The best results produced are highlighted.

Table 5
Average PV fault classification accuracy over seven FE for medium-scale PV array model

Feature Extraction	Average Accuracy of PV Fault Classification (%)	
	Training RF-CV algorithm	Testing RF-CV algorithm
P_{max}	<u>100.00</u>	97.33
M	<u>100.00</u>	79.33
STD	98.25	95.67
RMS	90.67	87.33
$Skew$	82.33	77.33
WL	98.08	<u>97.67</u>
AR	99.58	89.33

Based on the data shown, P_{max} and *M* features achieved perfect accuracy of 100.00% in training RF-CV algorithm stages. The features *STD*, *WL* and *AR* also demonstrated high average accuracy scores above 98% in the training work, with scores of 98.25%, 98.08% and 99.58%. However, the *RMS* and *Skew* features exhibited slightly lower accuracy. The *RMS* achieved scores of 90.67% in testing and 87.33% in training, while the *Skew* obtained scores of 82.33% in testing and 77.33% in training. Meanwhile, the average accuracy of the proposed combination FE is shown in Table 6. The best results produced are highlighted.

Table 6
Combination of feature extraction accuracy (Medium-scale PV model)

Feature Extraction	Average Accuracy of PV Fault Classification (%)	
	Training RF-CV Algorithm	Testing RF-CV Algorithm
$P_{max} + M + STD$	100.00	97.67
$P_{max} + M + STD + WL$	100.00	97.77
$P_{max} + M + STD + RMS$	100.00	98.00
$P_{max} + M + STD + AR$	100.00	98.33
$P_{max} + M + STD + AR + WL$	100.00	98.33
$P_{max} + M + STD + RMS + Skew$	100.00	99.33
$P_{max} + M + STD + RMS + Skew + WL$	100.00	99.00
$P_{max} + M + STD + RMS + Skew + AR$	100.00	99.00

Analysing the data provided in Table 6, which presents the average accuracy of PV fault classification for the proposed combination of FE in the medium-scale PV model, all combinations of feature extractions achieved a perfect accuracy of 100.00% during the training phase. In the testing phase, the combinations (P_{max}+*M*+*STD*+*RMS*+*Skew*), (P_{max}+*M*+*STD*+*RMS*+*Skew*+*WL*) and (P_{max}+*M*+*STD*+*RMS*+*Skew*+*AR*), achieved very good accuracy scores of 99.00% and above. The combination set of (P_{max}+*M*+*STD*), on the other hand, demonstrated the lowest accuracy of 97.67% during the testing phase. These results indicate that the addition of *RMS* and *Skew* features, along with the previously mentioned features of P_{max}, *M*, *STD* enhanced the fault classification accuracy in the medium-scale PV model.

This indicates that the classification model achieved a strong performance in correctly identifying instances across all fault types. The classification accuracy of each fault type produced from this training and testing RF-CV algorithm with the feature extraction set (P_{max}+*M*+*STD*+*RMS*+*Skew*) is described in detail in Table 7.

Table 7
Fault type classification accuracies (Medium-scale PV model)

RF-CV algorithm	Fault type classification accuracy (%)				
	GF	LLF	OCF	PS	Average
Training	100.00	100.00	100.0	100.0	100.0
Testing	100.00	97.33	100.00	100.00	99.33

4.3 Impact of Feature Extraction (FE) on Training and Testing with RF-CV Algorithm (Large-Scale PV Model)

Table 8 displays the average fault classification accuracy for a large-scale PV array model utilizing seven FE chosen in this research. The accuracy values are presented for both the training and testing simulations conducted using the RF-CV algorithm.

Table 8
Average PV fault classification accuracy over seven FE for large-scale PV array model

Feature Extraction	Average Accuracy of PV Fault Classification (%)	
	Training RF-CV algorithm	Testing RF-CV algorithm
P_{max}	97.92	96.00
M	99.25	77.67
STD	98.00	96.33
RMS	90.17	85.67
$Skew$	82.58	79.67
WL	98.08	98.00
AR	99.42	88.00

The data obtained shows that during the training phase, the RF-CV algorithm accomplished the most remarkable average accuracy of 99.42% in accurately classifying fault types by utilizing the AR. In addition, the features *STD*, *WL* and *M* also displayed impressive average accuracy scores exceeding 98% in the training process, with respective scores of 98.00%, 98.08% and 99.28%. However, the *RMS* and *Skew* features showed slightly lower accuracy, with scores of 90.17% and 82.58%, respectively.

In contrast, during the testing phase, the RF-CV algorithm attained the highest average accuracy of 98.00% in correctly classifying fault types by employing the WL FE method. The features *STD* and P_{max} also showed good average accuracy scores surpassing 96%, with corresponding scores of 96.00% and 96.32%. However, the *M* and *Skew* features demonstrated lower accuracy rates, achieving scores of 77.67% and 79.67%, respectively, which were below the 80%.

Then, the average accuracies of the proposed combination FE method are presented in Table 9, where the highest accuracy scores are highlighted.

Table 9
Combination set feature extraction accuracy (Large-scale PV model)

Feature Extraction	Average Accuracy of PV Fault Classification (%)	
	Training RF-CV Algorithm	Testing RF-CV Algorithm
$P_{max} + M + STD$	100.00	97.00
$P_{max} + M + STD + WL$	100.00	97.33
$P_{max} + M + STD + RMS$	100.00	97.33
$P_{max} + M + STD + AR$	100.00	97.67
$P_{max} + M + STD + AR + WL$	100.00	97.67
$P_{max} + M + STD + RMS + Skew$	100.00	98.33
$P_{max} + M + STD + RMS + Skew + WL$	100.00	98.67
$P_{max} + M + STD + RMS + Skew + AR$	100.00	98.00

It is noted that all combinations of feature extractions achieved a perfect accuracy of 100.00% during the training phase. In the testing phase, however, the combination feature set of (P_{max}+*M*+*STD*+*RMS*+*Skew*), (P_{max}+*M*+*STD*+*RMS*+*Skew*+*WL*) and (P_{max}+*M*+*STD*+*RMS*+*Skew*+*AR*), demonstrated good accuracy scores of 98.00% and above. The combination feature set of (P_{max}+*M*+*STD*), on the other hand, exhibited the lowest accuracy of 97.00% during the testing phase. These results indicate that the addition of *RMS* and *Skew* features, along with the previously mentioned features P_{max}, *M*, *STD* and *WL*, have significantly improved the fault classification accuracy in the large-scale PV model.

The classification accuracy of each fault type produced by this training and testing RF-CV algorithm with the feature extraction set (P_{max}+*M*+*STD*+*RMS*+*Skew*+*WL*), which achieved the highest average classification accuracy, is described in detail in Table 10.

Table 10
Fault type classification accuracies (Large-scale PV model)

RF-CV algorithm	Fault type classification accuracy (%)				
	GF	LLF	OCF	PS	Average
Training	100.00	100.00	100.00	100.00	100.00
Testing	100.00	96.00	98.67	100.00	98.67

From Table 10, the algorithm achieved a perfect fault type classification accuracy of 100% for all fault types, including GF, LLF, OCF and PS. Temporarily, during testing, the RF-CV algorithm demonstrated high accuracy as well. It correctly classified 100% of instances for GF and OCF fault types. For LLF, the algorithm achieved a slightly lower accuracy of 97.33%. Similarly, the algorithm achieved a perfect accuracy of 100% for the PS fault type. The average fault type classification accuracy for the testing dataset was 98.67%. Overall, the RF-CV algorithm exhibited excellent performance in accurately classifying fault types in the large-scale PV model, with high accuracies observed across all fault types during both training and testing phases.

5. Conclusions

This research developed the SFDC model as a smart fault diagnosis and classification model and implemented the RF-CV algorithm and an optimal FE combination. The SFDC model is developed from the PV array base model with minimal tuning to produce a small, medium and large-scale PV array model. This step is important for the SFDC model practicable to a multi-scale PV array model in real-world conditions. Furthermore, PV array models developed in this research have satisfactorily proven accurate in predicting their performance under both normal and faulty conditions. Other than that, the use of an optimal FE combination of *RMS*, *Skew*, P_{max}, *M*, *STD*, *WL* and *AR* significantly enhanced the fault classification accuracy of the SFDC model. By examining the most relevant features, the SFDC model achieved better accuracy for the training and testing of the RF-CV algorithm. All models (small, medium, large-scale) developed have achieved 100% accuracy for all fault types in training simulations. While in testing the algorithm, it effectively detected multiple fault types, especially OCF and PS, with a perfect score and slightly lower yet high accuracies for GF. These results indicate the robustness and reliability of the SFDC model in diagnosing and classifying faults in PV systems. The consistently lower accuracy for LLF across all models, however, highlights an area for potential improvement in future work.

Acknowledgement
This research was funded by grants from Universiti Malaysia Pahang Al-Sultan Abdullah with reference number PGRS2003189 and RDU223001.

References
[1] Ma, Raja Noor Farah Azura Raja, Nur Amira Syahirah Azmar, Mohd Syahriman Mohd Azmi, Nor Suriya Abd Karim, Nor Hafizah Md Husin, Norhisam Misron and Nida Sri Utami. "Sustainable Urban Planning: Criteria for Efficient Photovoltaic Electric Vehicle Charging Stations Deployment in Malaysia." *Semarak International Journal of Electronic System Engineering* 1, no. 1 (2024): 35-45. https://doi.org/10.37934/sijese.1.1.3545

[2] Adetona, Zacchaeus A., Joel Ogunyemi and Irmiya Bitrus. "Maintenance Management Regime for Off-Grid Solar PV Renewable Energy System in Nigeria." *European Journal of Engineering and Technology Research* 5, no. 11 (2020): 1376-1382. https://doi.org/10.24018/ejeng.2020.5.11.2233

[3] Saad, N. Md, M. Z. Sujod, M. I. M. Ridzuan, M. F. Abas, M. S. Jadin, M. S. Bakar and A. Z. Ahmad. "Solar irradiance uncertainty management based on Monte Carlo-beta probability density function: case in Malaysian tropical climate." *Bulletin of Electrical Engineering and Informatics* 8, no. 3 (2019): 1135-1143. https://doi.org/10.11591/eei.v8i3.1581

[4] Vaverková, Magdalena Daria, Jan Winkler, Dan Uldrijan, Paweł Ogrodnik, Tereza Vespalcová, Joanna Aleksiejuk-Gawron, Dana Adamcová and Eugeniusz Koda. "Fire hazard associated with different types of photovoltaic power plants: Effect of vegetation management." *Renewable and Sustainable Energy Reviews* 162 (2022): 112491. https://doi.org/10.1016/j.rser.2022.112491

[5] Al-Shetwi, Ali Q. and Muhamad Zahim Sujod. "Harmonic distortion and voltage imbalance study of photovoltaic power plant connected to the Malaysian grid." *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)* 10, no. 1-2 (2018): 1-6.

[6] Sujod, Muhamad Zahim and Istvan Erlich. "Harmonics and common mode voltage in a DFIG with two-level and three-level NPC converter using standard PWM techniques." In *IECON 2013-39th Annual Conference of the IEEE Industrial Electronics Society*, pp. 1650-1655. IEEE, 2013. https://doi.org/10.1109/IECON.2013.6699380

[7] Wu, Zuyu, Yihua Hu, Jennifer X. Wen, Fubao Zhou and Xianming Ye. "A review for solar panel fire accident prevention in large-scale PV applications." *IEEE Access* 8 (2020): 132466-132480. https://doi.org/10.1109/ACCESS.2020.3010212

[8] Fadhel, Siwar, Claude Delpha, Demba Diallo, I. Bahri, A. Migan, Mohamed Trabelsi and Mohamed Faouzi Mimouni. "PV shading fault detection and classification based on IV curve using principal component analysis: Application to isolated PV system." *Solar Energy* 179 (2019): 1-10. https://doi.org/10.1016/j.solener.2018.12.048

[9] Ghazali, Siti Nor Azlina Mohd and Muhamad Zahim Sujod. "A Comparative Analysis of Solar Photovoltaic Advanced Fault Detection and Monitoring Techniques." *Electrica* 23, no. 1 (2023).

[10] Ghazali, Siti Nor Azlina Mohd, Muhamad Zahim Sujod and Mohd Shawal Jadin. "Forensic of solar PV: A review of potential faults and maintenance strategies." In *2021 International Conference on Engineering and Emerging Technologies (ICEET)*, pp. 1-6. IEEE, 2021. https://doi.org/10.1109/ICEET53442.2021.9659624

[11] Yang, Nien-Che and Harun Ismail. "Robust intelligent learning algorithm using random forest and modified-independent component analysis for PV fault detection: In case of imbalanced data." *IEEE Access* 10 (2022): 41119-41130. https://doi.org/10.1109/ACCESS.2022.3166477

[12] Din, Roshidi, Nuramalina Mohammad Na'in, Sunariya Utama, Muhaimen Hadi and Alaa Jabbar Qasim Almaliki. "Innovative Machine Learning Applications in Non-Revenue Water Management: Challenges and Future Solution." *Semarak International Journal of Machine Learning* 1, no. 1 (2024): 1-10. https://doi.org/10.37934/sijml.1.1.110

[13] Elkholy, A. and A. A. Abou El-Ela. "Optimal parameters estimation and modelling of photovoltaic modules using analytical method." *Heliyon* 5, no. 7 (2019). https://doi.org/10.1016/j.heliyon.2019.e02137

[14] Bacha, Marah and Amel Terki. "Diagnosis algorithm and detection faults based on fuzzy logic for PV panel." *Materials Today: Proceedings* 51 (2022): 2131-2138. https://doi.org/10.1016/j.matpr.2021.12.490

[15] Akhtar, Shamim, Muhamad Zahim Bin Sujod and Syed Sajjad Hussain Rizvi. "An intelligent data-driven approach for electrical energy load management using machine learning algorithms." *Energies* 15, no. 15 (2022): 5742. https://doi.org/10.3390/en15155742

[16] Murtaza, Ali Faisal, Mohsin Bilal, Riaz Ahmad and Hadeed Ahmed Sher. "A circuit analysis-based fault finding algorithm for photovoltaic array under L–L/L–G faults." *IEEE Journal of Emerging and Selected Topics in Power Electronics* 8, no. 3 (2019): 3067-3076. https://doi.org/10.1109/JESTPE.2019.2904656

[17] Al-Shetwi, Ali Q. and Muhamad Zahim Sujod. "Modeling and simulation of photovoltaic module with enhanced perturb and observe MPPT algorithm using Matlab/Simulink." *sat* 25 (2006): 25.

[18] Dhibi, Khaled, Majdi Mansouri, Kais Bouzrara, Hazem Nounou and Mohamed Nounou. "An enhanced ensemble learning-based fault detection and diagnosis for grid-connected PV systems." *IEEE Access* 9 (2021): 155622-155633. https://doi.org/10.1109/ACCESS.2021.3128749

[19] Ghazali, S. N. A. M. and Muhamad Zahim Sujod. "A multi-scale dual-stage model for PV array fault detection, classification and monitoring technique." *International Journal of Applied Power Engineering (IJAPE)* 11, no. 2 (2022): 134. https://doi.org/10.11591/ijape.v11.i2.pp134-144

[20] Ghazali, Siti Nor Azlina M., Muhamad Zahim Sujod and Mohd Shawal Jadin. "A multi-scale smart fault diagnosis model based on waveform length and autoregressive analysis for PV system maintenance strategies." *Chinese Journal of Electrical Engineering* (2023). https://doi.org/10.23919/CJEE.2023.000023

[21] Ishak, Norshahidatul Hasana, Nur Farahin Johari, Siti Nuramalina Johari, Hazrati Zaini, Abdul Hasib Sadiqin Adam Shukri and Aqil Afifi Ahmad Senusi. "Learning and Training of Coal-Fired Energy using Augmented Reality (C-FAR)." *Journal of Advanced Research in Computing and Applications* 17, no. 1 (2019): 20-24.

[22] Zhang, Jingwei, Yongjie Liu, Yuanliang Li, Kun Ding, Li Feng, Xihui Chen, Xiang Chen and Jiabing Wu. "A reinforcement learning based approach for on-line adaptive parameter extraction of photovoltaic array models." *Energy Conversion and Management* 214 (2020): 112875. https://doi.org/10.1016/j.enconman.2020.112875

[23] Andrianajaina, T., DT RAZAFIMAHEFA, Nicolas Héraud and Eric Jean Roy Sambatra. "ANOVA fault detection method in photovoltaic system." In *2020 7th International Conference on Control, Decision and Information Technologies (CoDIT)*, vol. 1, pp. 37-41. IEEE, 2020. https://doi.org/10.1109/CoDIT49905.2020.9263843
[24] Madeti, Siva Ramakrishna and S. N. Singh. "Modeling of PV system based on experimental data for fault detection using kNN method." *Solar Energy* 173 (2018): 139-151. https://doi.org/10.1016/j.solener.2018.07.038
[25] Ray, Prakash K., Asit Mohanty, Basanta K. Panigrahi and Pravat K. Rout. "Modified wavelet transform based fault analysis in a solar photovoltaic system." *Optik* 168 (2018): 754-763. https://doi.org/10.1016/j.ijleo.2018.03.131
[26] Nasir, Muhammad Nabihan Arif, Nolia Harudin, Farah Elida Selamat and Zulkifli Marlah. "Development Of Enhanced Taguchi's T-Method Model in Predicting Energy Demand." *Semarak Engineering Journal* 1, no. 1 (2023): 34-45.