Received: 21 July 2023 | *Revised: 2 September 2023* | *Accepted: 22 December 2023*
DOI: 10.37917/ijeee.21.1.12 *Early View* | *June 2025*

Open Access

Iraqi Journal for Electrical and Electronic Engineering
Original Article

Enhancing PV Fault Detection Using Machine Learning: Insights from a Simulated PV System

Halah Sabah Muttashar*, Amina Mahmoud Shakir
Department of Electronic and Communications Engineering, Al-Nahrain University, Iraq

Correspondance
*Halaa Sabah Muttashar
Department of Electronic and Communications Engineering,
Al-Nahrain University, Iraq
Email: st.hala.sabah.1@nahrainuniv.edu.iq

Abstract
Recently, numerous researches have emphasized the importance of professional inspection and repair in case of suspected faults in Photovoltaic (PV) systems. By leveraging electrical and environmental features, many machine learning models can provide valuable insights into the operational status of PV systems. In this study, different machine learning models for PV fault detection using a simulated 0.25 ***MW PV power system were developed and evaluated. The training and testing datasets encompassed normal operation and various fault scenarios, including string-to-string, on-string, and string-to-ground faults. Multiple electrical and environmental variables were measured and exploited as features, such as current, voltage, power, temperature, and irradiance. Four algorithms (Tree, LDA, SVM, and ANN) were tested using 5-fold cross-validation to identify errors in the PV system. The performance evaluation of the models revealed promising results, with all algorithms demonstrating high accuracy. The Tree and LDA algorithms exhibited the best performance, achieving accuracies of*** 99.544% ***on the training data and*** 98.058% ***on the testing data. LDA achieved perfect accuracy*** (100%) ***on the testing data, while SVM and ANN achieved*** 95.145% ***and*** 89.320% ***accuracy, respectively. These findings underscore the potential of machine learning algorithms in accurately detecting and classifying various types of PV faults. .***

Keywords
PV Fault Detection; Standalone PV System; Machine Learning; Classifiers; Training and Testing Accuracy; Confusion Matrix.

I. Introduction

Photovoltaic (PV) systems are becoming increasingly popular as a source of renewable energy PV systems, like any other electrical system, are prone to defects that may reduce their efficiency and lifespan. Detecting and addressing these faults is crucial for maintaining the performance and reliability of PV systems. In recent years, many strategies for fault detection and diagnosis in PV systems have been proposed. These algorithms employ several methodologies, including machine learning, optimization, and signal processing, to discover and diagnose problems in PV systems [1–5]. Artificial Neural Networks (ANNs) [6]. Support Vector Machines (SVMs) [7], and Deep Learning [8] have been shown to be useful in identifying and diagnosing defects in PV systems. To understand the patterns associated with each kind of operation, ANNs are trained using a dataset of normal and problematic PV system operations. Once trained, the ANN may utilize the learned patterns to identify new data samples as normal or erroneous. Support Vector Machines (SVMs) are another type of machine learning algorithm that can be used for fault detection and diagnosis in PV systems. SVMs function by locating the hyperplane in a high-dimensional feature space that optimally distinguishes normal and erroneous data points [9]. After locating the hyperplane, the applied data may be classed as normal or erroneous depending on which side of the hyperplane it falls on. Deep Learning uses multiple-layer artificial neural networks to learn complicated patterns in data. A decision Tree is a type of machine learning algorithm that can be used for fault detection and diagnosis in PV systems. Decision Trees work by recursively partitioning the data into subsets based on the values of the input features. At each node of the In a decision tree, a decision is made based on the input features to determine which subset of the data to partition next. Once the tree is built, new data can be classified as normal or faulty by traversing the tree from the root to a leaf node. Decision trees are useful in identifying and diagnosing issues in PV systems [10] with a good accuracy of 95.5%.

Linear Discriminant Analysis (LDA) is useful in identifying and diagnosing defects in PV systems. Two researchers employed LDA [11, 12]to categorize PV system data as normal or problematic based on characteristics such as voltage, current, and power. The findings revealed that LDA detected and diagnosed issues in the PV system with excellent accuracy.

Optimization methods like Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) [13] may be used to increase the efficiency of a defect detection program by optimizing its parameters according to the constraint functions and values. GAs simulate the process of natural selection to enhance the parameters of a fault detection algorithm and hence increase its performance in the setting of problem detection and diagnosis. PSO simulates the behavior of a swarm of particles as they travel across a search space in pursuit of the best solution. One work [9] employed a PSO for extracting some features while the SVM was used to improve the parameters and defect detection in a PV system.

To identify and diagnose issues in PV systems, signal processing methods such as Wavelet Transform (WT) and Principal Component Analysis (PCA) may be utilized. The WT decomposes a signal into a series of wavelets with varying frequencies and temporal scales. The wavelets may then be studied to discover and diagnose signal flaws. PCA works by converting the data into a new collection of uncorrelated variables that are ranked by their relevance in explaining the data's variance. After that, the modified data may be evaluated to find and diagnose errors. In one research, WT was utilized to identify and diagnose defects in a PV system [14]. PCA was used to identify and diagnose defects in a PV system in another work [15]. Deep learning with ensemble techniques was used to identify and diagnose defects in a PV system. The method detected errors with an accuracy of 98.5% [16].

In this work, a proposed 0.25 MW standalone PV power system is investigated and tested under different environmental conditions. As a result, a dataset with varying features is generated, and then several machine learning algorithms for fault detection and diagnosis including Decision Tree (DT), Linear Discriminant Analysis (LDA), Support Vector Machine (SVM) with linear kernel, and Artificial Neural Network (ANN) with simple structure are applied. Each of these algorithms has its strengths and weaknesses, and the choice of algorithm will depend on the application's specific requirements.

II. Standalone Mode Photovoltaic System

A standalone PV system as shown in Fig. 1 is a self-sufficient system that generates and stores electricity without being connected to the grid. It consists of several components that work together to generate and store electricity [17]. PV modules, charge controller, battery bank, inverter, and load are the primary components of a standalone PV system. PV modules are panels that turn sunlight into energy and are composed of several solar cells that are linked together. The charge controller manages the amount of current given to the battery bank, avoiding overcharging and undercharging, which may harm the batteries. The battery bank is where the electricity generated by the PV modules is stored, and it is made up of multiple batteries that are connected. The inverter converts the DC electricity stored in the battery bank into AC electricity that can be used by household appliances. The load is the electrical load that is powered by the PV system, and it can be any electrical device or appliance that requires electricity. Faults can occur in any of these components, which can reduce the efficiency and lifespan of the PV system. It is important to regularly inspect and maintain the components of a standalone PV system to prevent faults from occurring.

III. The Proposed Microgrid Architecture

Fig. 1 illustrates the proposed configuration for the microgrid, which includes a photovoltaic system connected to the direct current link using a unidirectional boost converter, a battery storage system connected to a bidirectional buck-boost converter to maintain system stability across different disturbance situations, a two-level inverter connected to the LC filter, a variable load for the test system connected at PCC, and the entire system connected to the main load using an inverter. The solar plant is a Sunpower SPR-400E model, and it has 12 panels connected in series and 52 panels connected in parallel to generate a total power of 250 KW. Each panel has a capacity of 400 W and 90 cells, and the plant generates a total of 250 KW. Tables I and II include a listing of the characteristics and measurements of the PV panels that were used for this study. Between 200 and 250 kW of electricity are produced by the solar power plant as a whole to meet the demand load. All simulation investigations were carried out with the help of the MATLAB/Simulink program.

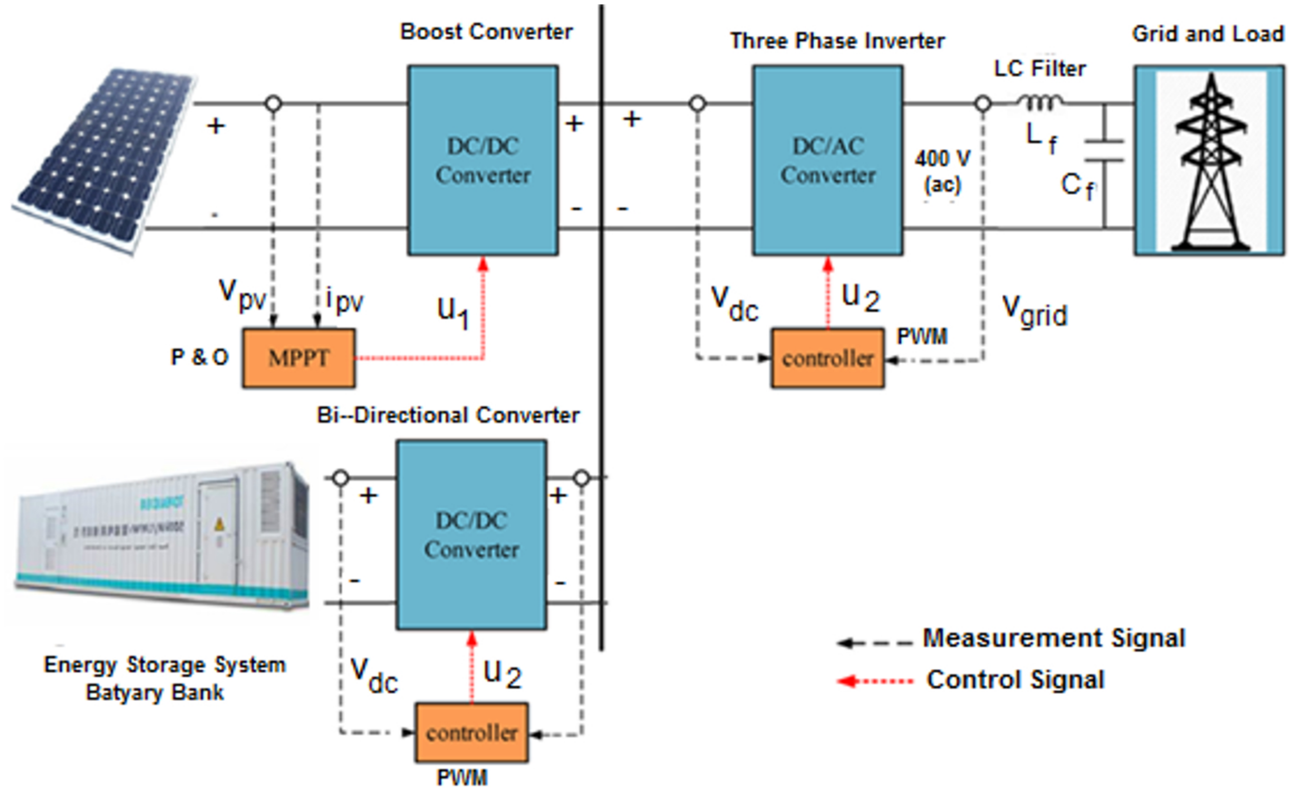


Fig. 1. The standalone/grid PV system block diagram.

TABLE I.
THE TECHNICAL SPECIFICATIONS OF THE PROPOSED PV SYSTEM

Symbol	Description	Value
V_{OC}	Open circuit voltage	85.3 V
I_{SC}	Short circuit current	5.87 A
I_{mpp}	Current at the maximum power point	5.49 A
V_{mpp}	Voltage at the maximum power point	72.9 V
MPP	Max power point	400 W
N_s	Number of series panels	12
N_v	Number of parallel panels	52
P_{total}	Total power of solar power plant	250 kW

TABLE II.
THE DETERMINED PARAMETERS OF THE SUGGESTED SYSTEM

Parameters	Value	Parameters	Value
Rated power	200 kW	Boost capacitor	3 mF
Grid line voltage	400 V	Boost inductance	0.8 mH
DC-Voltage	800 V	DC-link capacitor	5 mF
Grid frequency	50 Hz	Filter capacitor	40 μF
Switching frequency	10,000 Hz	Bi-directional inductance	0.3 mH

IV. DATASET PREPARATION

In order to generate training and testing datasets of PV defect scenarios, a simulated 250kW PV power system was developed. Normal operation (fault-free state with F0) is specified and three different fault kinds as can be seen in Fig. 2 were employed. The predefined sets of faults consist of string-to-string fault (F1), on-string fault (F2), and string-to-ground fault (F3), respectively.

Training and testing datasets were built on the DC side of the PV system. The normal case and the different types of photovoltaic (PV) faults are listed as:

1. **Fault-free (F0)**: This is the ideal scenario where the PV system is operating normally without any faults or issues.

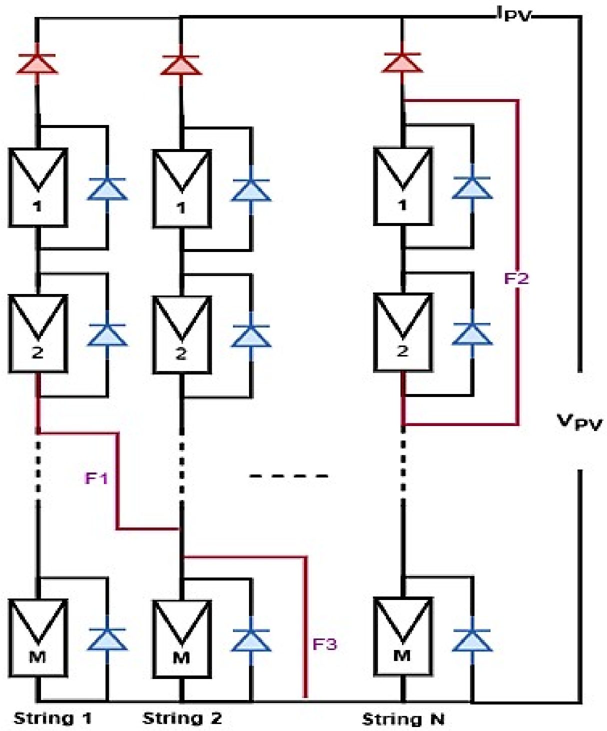


Fig. 2. A sample of M×N PV array with three different faults illustration.

2. **String-to-string fault (F1)**: This type of fault occurs when there is a problem with the wiring between two strings of PV modules. This can happen due to a loose connection, damaged wiring, or a faulty connector. As a result, the affected string may not produce power, or its output may be reduced.

3. **On-string fault (F2)**: When one or more PV modules within a string have a malfunction, this is known as an on-string fault. A broken module, a malfunctioning bypass diode, or improper shading might all be to blame. The overall string performance may suffer if the afflicted module(s) either do not generate power or provide much less power than usual.

4. **String-to-ground fault (F3)**: When there is an issue with the wiring connecting a series of PV modules to the ground, this fault will occur. A broken grounding wire or an imperfect grounding electrode might cause this. There may be safety problems owing to the possibility of electrical shock, and there may be no or limited power production from the damaged string. These are only some of the potential PV defects that might arise; additional problems may also affect the efficiency of a PV system.

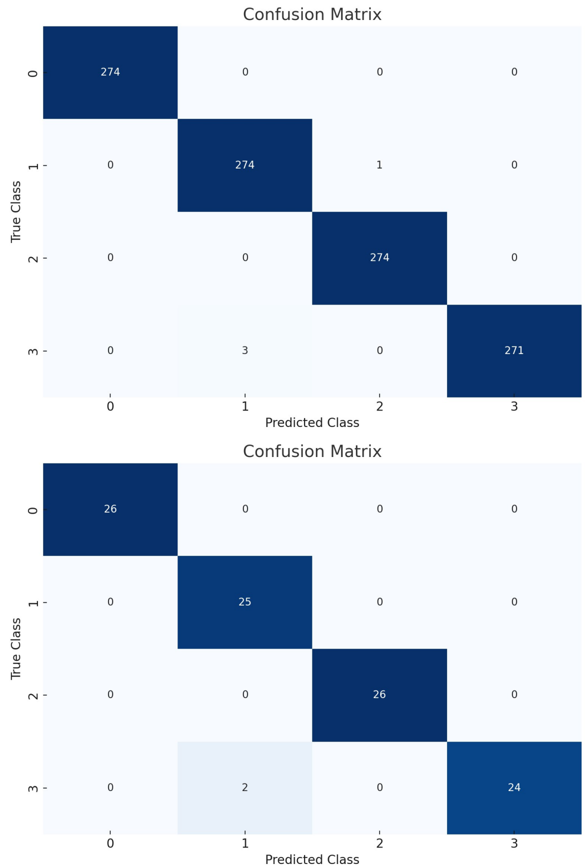
The dataset includes different electrical features such as minimum, maximum, average, and ranges of the currents, voltages, and powers. In addition, different environmental features like temperature and irradiations are added. Temperature (T) may be in the range of 10°C to 45°C, with a 5°C resolution. Light irradiation (G) has values from $100\,\mathrm{W/m^2}$ to $1000\,\mathrm{W/m^2}$ with a $50\,\mathrm{W/m^2}$ step. Our suggested models' accuracy in defect identification was evaluated using the 5-fold cross-validation approach, and its precision was calculated using an unbiased estimate. Each dataset was randomly split into a training set (consisting of 90% of the data) and a test set (consisting of 10% of the data). As a result, a balanced dataset with the same number of samples from each category was included in each set. The overall performance was obtained by determining the average for all 5 trials [18–23]. The dataset consists of 1200 instances, each with 24 features and one column for the fault classes; training data consists of 1097×24 while testing data consists of the remaining 103×24. The simulations will account for most of the I-V characteristics curve of the PV array under varying environmental conditions. The distribution of PV simulation datasets with all failure types is shown in Table III.

TABLE III.
THE PV SIMULATION DATASET DISTRIBUTION WITH VARIOUS FAULT TYPES

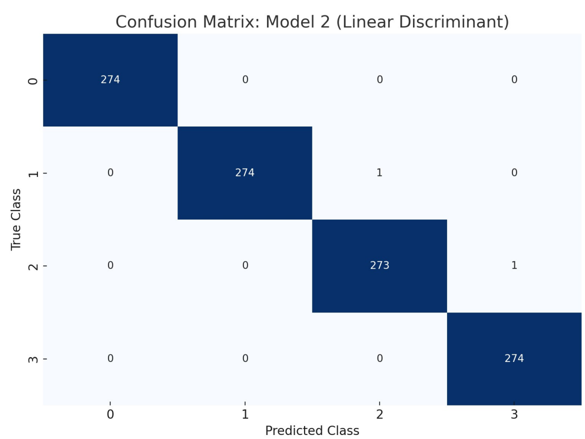
Fault Type	Nominal or class	Size
Fault free	F0	300×25
String-to-string fault	F1	300×25
On-string fault	F2	300×25
String-to-ground	F3	300×25
Total data = 1200×25, training part = 1097×25, testing part = 103×25		

V. EXPERIMENTAL RESULTS AND DISCUSSIONS

The data set is normalized first using Z-Score normalization (standardization), then applied to four classifiers (DT, LDA, SVM, and ANN), the training and testing confusion resultant matrices obtained from these four models are detailed in Fig. 3-a, b, c, and d. Table IV explains the performance metrics including the model's accuracy on the validation and test sets, as well as the training time for each model. It has arisen from the depths of data analysis in the area of machine learning [24–26], displaying the performance of four distinct models on a dataset. The dataset, "xtrain," has 1097 observations with a unique response variable labeled 0, 1, 2, and 3. The models were tested on a different dataset named "xtest," which has 103 observations, and were assessed using 5-fold

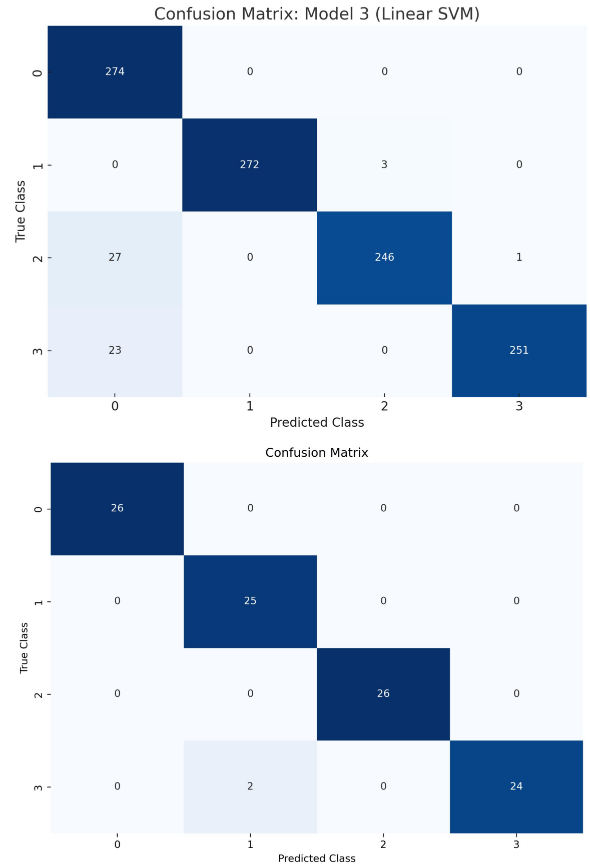


(a) DT train and test confusion matrices.

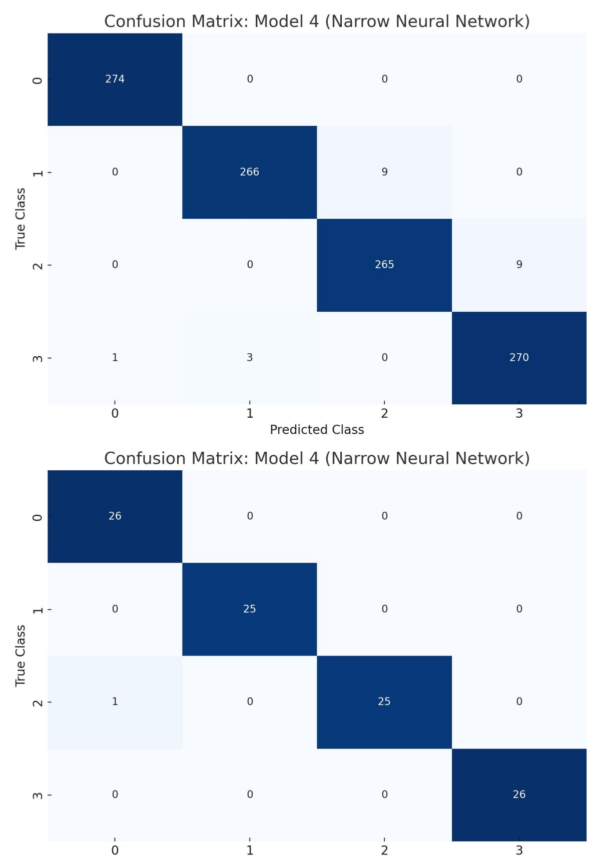


(b) LDA train and test confusion matrices

Fig. 3. Confusion matrices for DT, LDA, SVM, and ANN models



(c) SVM train and test confusion matrices.



(d) ANN train and test confusion matrices

cross-validation.
The training and testing accuracies are illustrated as in Fig. 4, while the training time is shown in Fig. 5. Model 2 (LDA) performed the best, attaining 99.73% and 100% accuracy on both the validation and test sets, respectively. Model 1 (decision tree) performed well as well, with 99.54% accuracy on the validation set and 98.06% accuracy on the test set. Model 3 (SVM) struggled to keep up, scoring 95.17% on the validation set and 95.15% on the test set. Model 4 (ANN) achieved 96.72% accuracy on the validation set but had the lowest accuracy on the test set owing to employing the simplest ANN structure, at 89.32% accuracy. It was also the model with the longest training time, coming in at 11.09 seconds.
While our study demonstrated strong accuracy in fault detection, it is noteworthy that these findings align with previous research efforts [27–30] in the domain of PV system fault detection. These previous studies have also emphasized the potential of machine learning models to accurately identify and classify different types of PV faults. Our results not only reaffirm these prior observations but also extend them by evaluating multiple machine learning algorithms on a comprehensive dataset.

TABLE IV.
ALL CLASSIFIERS' PERFORMANCES

Model Number	Model Type	% Train Accuracy	% Test Accuracy	Training Time (s)
1	Tree	99.544	98.058	3.1968
2	LDA	99.726	100	2.672
3	SVM	95.168	95.145	2.772
4	ANN	96.718	89.320	11.091
Training Data = 1097 observations, Test Data = 103 observations				

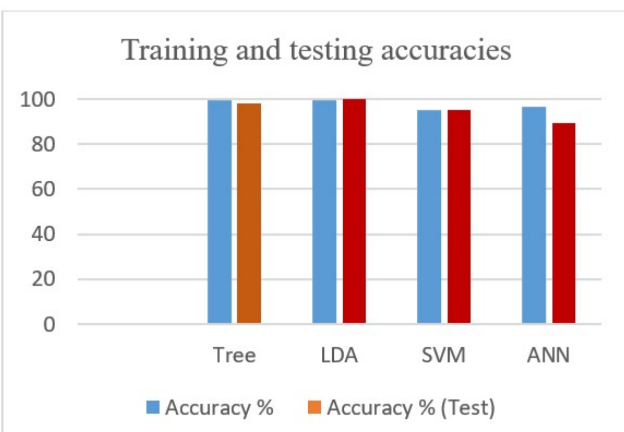


Fig. 4. The training and testing accuracies of all proposed classifiers.

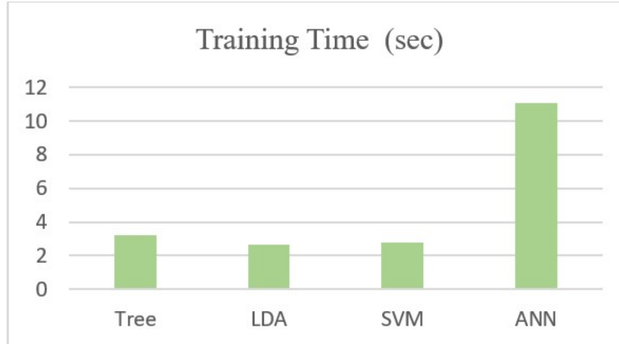


Fig. 5. The training time of all proposed classifiers.

VI. CONCLUSION

In this study, the development and evaluation of different nature machine learning models for PV fault detection using a simulated 0.25 MW PV power system is presented. By utilizing electrical and environmental features, these models can provide valuable insights into the operational status of PV systems. The datasets encompassed different fault types (string-to-string, on-string, and string-to-ground) as well as normal operation, capturing a range of electrical and environmental features. The performance of four algorithms (Tree, LDA, SVM, and ANN) was assessed using 5-fold cross-validation. The results demonstrated the effectiveness of the employed algorithms in detecting PV faults, with high accuracy observed for all models. DT or Tree algorithm exhibited the highest performance, achieving 99.544% accuracy on the training data and 98.058% accuracy on the testing data. LDA achieved perfect accuracy (100%) on the testing data, while SVM and ANN achieved 95.145% and 89.320% accuracy, respectively. These findings indicate the potential of machine learning algorithms in accurately identifying and classifying different types of PV faults.
Future works can suggest the following:

- **Explore additional fault types:** Investigate more complex faults such as partial shading, module degradation, or inverter malfunctions to improve the models' fault detection capabilities.
- **Validate with real-world data:** Verify the performance of the models using data from operational PV systems to assess their effectiveness in practical scenarios.
- **Feature selection and optimization:** Utilize feature selection techniques and optimization algorithms to improve model efficiency and accuracy.
- **Ensemble models:** Investigate the use of ensemble models to combine predictions from multiple algorithms, enhancing fault detection reliability.

- **Online fault detection:** Develop algorithms for real-time monitoring of PV systems, enabling proactive maintenance and minimizing system downtime.
- **Incorporate external data:** Integrate weather forecasts or historical data to account for environmental variations and improve fault detection accuracy.

Conflict of Interest

The authors have no conflict of relevant interest to this article

References

[1] B. Aljafari, "Reliability investigation of long photovoltaic string generators under electrical fault scenarios," in *Proceedings of the 2022 International Conference on Electrical and Computing Technologies and Applications (ICECTA)*, pp. 9–14, 2022.

[2] W. Zhang, H. Zhang, J. Liu, K. Li, D. Yang, and H. Tian, "Weather prediction with multiclass support vector machines in the fault detection of photovoltaic system," *IEEE/CAA Journal of Automatica Sinica*, vol. 4, pp. 520–525, 2017.

[3] A. Jäger-Waldau, J. Donoso, I. Kaizuka, G. Masson, and E. Bosch, "Snapshot 2022—iea-pvps," *EPJ Photovoltaics*, vol. 13, p. 9, 2022.

[4] W. Fenz, S. Thumfart, R. Yatchak, H. Roitner, and B. Hofer, "Detection of arc faults in pv systems using compressed sensing," *IEEE Journal of Photovoltaics*, vol. 10, pp. 676–684, 2020.

[5] R. Fazai, K. Abodayeh, M. Mansouri, M. Trabelsi, H. Nounou, M. Nounou, *et al.*, "Machine learning-based statistical testing hypothesis for fault detection in photovoltaic systems," *Solar Energy*, vol. 190, pp. 405–413, 2019.

[6] B. Li, C. Delpha, D. Diallo, and A. Migan-Dubois, "Application of artificial neural networks to photovoltaic fault detection and diagnosis: A review," *Renewable and Sustainable Energy Reviews*, vol. 138, p. 110512, 2021.

[7] J. Wang, D. Gao, S. Zhu, S. Wang, and H. Liu, "Fault diagnosis method of photovoltaic array based on support vector machine," *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, vol. 45, pp. 5380–5395, 2023.

[8] M. Pa, M. Uddin, and A. Kazemi, "A fault detection scheme utilizing convolutional neural network for pv solar panels with high accuracy," in *Proceedings of the 2022 IEEE 1st Industrial Electronics Society Annual On-Line Conference (ONCON)*, pp. 1–5, 2022.

[9] S. A. Memon, Q. Javed, W.-G. Kim, Z. Mahmood, U. Khan, and M. Shahzad, "A machine-learning-based robust classification method for pv panel faults," *Sensors*, vol. 22, p. 8515, 2022.

[10] A. Abid, M. T. Khan, and J. Iqbal, "A review on fault detection and diagnosis techniques: basics and beyond," *Artificial Intelligence Review*, vol. 54, pp. 3639–3664, 2021.

[11] A. Mellit, G. M. Tina, and S. A. Kalogirou, "Fault detection and diagnosis methods for photovoltaic systems: A review," *Renewable and Sustainable Energy Reviews*, vol. 91, pp. 1–17, 2018.

[12] T. Pei and X. Hao, "A fault detection method for photovoltaic systems based on voltage and current observation and evaluation," *Energies*, vol. 12, p. 1712, 2019.

[13] V. S. B. Kurukuru, A. Haque, M. A. Khan, S. Sahoo, A. Malik, and F. Blaabjerg, "A review on artificial intelligence applications for grid-connected solar photovoltaic systems," *Energies*, vol. 14, p. 4690, 2021.

[14] C. He, L. Mu, and Y. Wang, "The detection of parallel arc fault in photovoltaic systems based on a mixed criterion," *IEEE Journal of Photovoltaics*, vol. 7, pp. 1717–1724, 2017.

[15] T. G. Amaral, V. F. Pires, and A. J. Pires, "Fault detection in pv tracking systems using an image processing algorithm based on pca," *Energies*, vol. 14, p. 7278, 2021.

[16] E. Lodhi, F.-Y. Wang, G. Xiong, L. Zhu, T. S. Tamir, W. U. Rehman, *et al.*, "A novel deep stack-based ensemble learning approach for fault detection and classification in photovoltaic arrays," *Remote Sensing*, vol. 15, p. 1277, 2023.

[17] T. Khatib, I. A. Ibrahim, and A. Mohamed, "A review on sizing methodologies of photovoltaic array and storage battery in a standalone photovoltaic system," *Energy Conversion and Management*, vol. 120, pp. 430–448, 2016.

[18] I. M. Moreno-Garcia, E. J. Palacios-Garcia, V. Pallares-Lopez, I. Santiago, M. J. Gonzalez-Redondo, M. Varo-Martinez, *et al.*, "Real-time monitoring system for a utility-scale photovoltaic power plant," *Sensors*, vol. 16, p. 770, 2016.

[19] S. S. Ghoneim, A. E. Rashed, and N. I. Elkalashy, "Fault detection algorithms for achieving service continuity in photovoltaic farms," *Intelligent Automation and Soft Computing*, vol. 30, pp. 467–479, 2021.

[20] A. Triki-Lahiani, A. B. B. Abdelghani, and I. Slama-Belkhodja, "Fault detection and monitoring systems for photovoltaic installations: A review," *Renewable and Sustainable Energy Reviews*, vol. 82, pp. 2680–2692, 2018.

[21] Z. Yi and A. H. Etemadi, "Fault detection for photovoltaic systems based on multi-resolution signal decomposition and fuzzy inference systems," *IEEE Transactions on Smart Grid*, vol. 8, pp. 1274–1283, 2016.

[22] M. H. Ali, A. Rabhi, A. E. Hajjaji, and G. M. Tina, "Real time fault detection in photovoltaic systems," *Energy Procedia*, vol. 111, pp. 914–923, 2017.

[23] R. F. Colmenares-Quintero, E. R. Rojas-Martinez, F. Macho-Hernantes, K. E. Stansfield, and J. C. Colmenares-Quintero, "Methodology for automatic fault detection in photovoltaic arrays from artificial neural networks," *Cogent Engineering*, vol. 8, p. 1981520, 2021.

[24] B. K. Karmakar and A. K. Pradhan, "Detection and classification of faults in solar pv array using thevenin equivalent resistance," *IEEE Journal of Photovoltaics*, vol. 10, pp. 644–654, 2020.

[25] H. Al-Zubaidi, M. A. Shehab, and A. Al-Gizi, "Electrical fault diagnosis of solar pv array using machine learning techniques," in *Proceedings of the 2023 13th International Symposium on Advanced Topics in Electrical Engineering (ATEE)*, pp. 1–4, 2023.

[26] E. Garoudja, F. Harrou, Y. Sun, K. Kara, A. Chouder, and S. Silvestre, "Statistical fault detection in photovoltaic systems," *Solar Energy*, vol. 150, pp. 485–499, 2017.

[27] N. V. Sridharan and V. Sugumaran, "Visual fault detection in photovoltaic modules using decision tree algorithms with deep learning features," *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, pp. 1–17, 2021.

[28] N. C. Yang and H. Ismail, "Voting-based ensemble learning algorithm for fault detection in photovoltaic systems under different weather conditions," *Mathematics*, vol. 10, no. 2, p. 285, 2022.

[29] K. H. Cho, H. C. Jo, E. S. Kim, H. A. Park, and J. H. Park, "Failure diagnosis method of photovoltaic generator using support vector machine," *Journal of Electrical Engineering & Technology*, vol. 15, pp. 1669–1680, 2020.

[30] S. Rao, A. Spanias, and C. Tepedelenlioglu, "Solar array fault detection using neural networks," in *Proceedings of the 2019 IEEE International Conference on Industrial Cyber Physical Systems (ICPS)*, pp. 196–200, IEEE, May 2019.