Rev. Roum. Sci. Techn.– Électrotechn. et Énerg.
Vol. **70**, *4*, pp. 513–518, Bucarest, 2025

FAULT DETECTION AND DIAGNOSIS IN PHOTOVOLTAIC POWER SYSTEMS USING FISHER RANDOM MATRIX APPROACH

AHMED SAIDI[1], DRAOUI ABDELGHANI[2], ABDELOUAHED TOUHAMI[3]

Keywords: Fault detection and diagnosis (FDD); Photovoltaic (PV) systems; Fisher random matrix theory (RMT); Real-time monitoring; Multidimensional data analysis.

This paper proposes a novel methodology for fault detection and diagnosis (FDD) in photovoltaic (PV) systems that combines Fisher's linear discriminant (FLD) with the Mahalanobis distance. The approach utilizes FLD to reduce the dimensionality of operational data while maximising the separation between healthy and faulty states. The Mahalanobis distance is then used to detect anomalies by accounting for correlations among variables such as voltage, current, and temperature. The validity of the method was established using real-world data from a 20 MWp PV plant in Algeria. The results obtained demonstrate the efficacy of the proposed method in classifying various faults, including open-circuit, shading, and short-circuit faults. The approach demonstrates substantial improvements in detection accuracy, efficiency, and false-alarm reduction compared to conventional methodologies. The proposed FDD solution is robust and scalable, rendering it ideal for real-time monitoring of large-scale PV systems.

1. INTRODUCTION

Photovoltaic (PV) systems are considered pivotal to the global energy transition; however, concerns regarding their long-term viability have been raised. These concerns include various failures, such as shading, open circuits, short circuits, and critical ground faults [1–5]. These faults have been shown to have a significant impact on energy yield and pose a considerable safety risk. This emphasizes the critical importance of research into fault detection and diagnosis (FDD) for maintaining the system's reliability and efficiency [3].

The evolution of FDD has progressed through distinct methodological stages. Early approaches relied heavily on manual inspection (visual or infrared imaging), a method limited to superficial defects and not suitable for large installations [6–8]. The initial significant progression was marked by the advent of electrical-based methodologies, which predominantly concentrated on anomalies in the current-voltage (I-V) and power-voltage (P-V) characteristics [9,10]. Whilst these model-based techniques are indeed effective, they are also highly sensitive to fluctuations in operational conditions (irradiance and temperature), which frequently result in false alarms.

To address this sensitivity, researchers used statistical methods – such as Principal Component Analysis (PCA) and t-tests –to better filter environmental noise and establish operational baselines [11–13]. Concurrently, the field embraced machine learning (ML) techniques such as artificial neural networks (ANNs) and support vector machines (SVMs) to model complex PV system behaviour and improve fault classification accuracy [14–16]. More recently, the field of deep learning (DL) has witnessed significant advances, particularly in convolutional neural networks (CNNs) and long short-term memory (LSTM) networks. These architectures have demonstrated remarkable efficacy in enhancing accuracy, often in conjunction with advanced data types such as infrared thermography [17–19].

Notwithstanding these advances, a considerable challenge remains: the development of an FDD methodology that offers high accuracy, noise immunity, and precise fault isolation across all operating conditions. Hybrid and intelligent systems show promise, but many still struggle with complex multivariate correlations in high-dimensional real-time data [19–21].

The present paper addresses this gap by presenting an innovative FDD methodology based on Fisher random matrix theory (RMT). The exploitation of multivariate statistical correlations within high-dimensional operational data is a key element of the methodology, enabling highly accurate, real-time fault detection. The employment of Fisher RMT confers this technique a distinct advantage in enhancing fault isolation and diagnosis while ensuring minimal sensitivity to environmental noise. The validity of the proposed method is demonstrated through the utilisation of real PV data, which shows substantial improvements in detection speed, accuracy, and false alarm mitigation compared to conventional I-V curves and extant FDD techniques.

The remainder of this paper is organized as follows: Section 2 details the proposed methodology, including the theoretical background of Fisher random matrix and Mahalanobis distance. Section 3 describes the experimental setup and the validation dataset. Section 4 presents and discusses the results, demonstrating the efficacy of the proposed approach. Finally, Section 5 concludes the paper and outlines potential directions for future work.

2. MATERIALS AND METHODS

In the context of fault detection and diagnosis systems, particularly in photovoltaic (PV) systems, the use of a robust method is imperative for effectively distinguishing between normal and faulty operation. A methodology found to be effective involves using the Fisher random matrix in conjunction with the Mahalanobis distance, thereby facilitating the establishment of an effective threshold for fault detection.

2.1 FISHER RANDOM MATRIX

The Fisher random matrix is derived from Fisher's linear discriminant analysis (LDA), a method widely used for dimensionality reduction and classification. The objective of LDA is to project high-dimensional data onto a lower-dimensional space while maximizing the separation between different classes, such as healthy and faulty states in a PV system. The Fisher random matrix is employed to transform the data in a manner that accentuates the disparities between these classes.

[1] Department of Electrical Engineering, Laboratory of Sustainable Development and Computer Science, Ahmed Draia University, Adrar 01000, Algeria.
[2] Laboratory of environmental and energy systems (LSEE) University center Ali Kafi Tindouf 37000, Algeria.
[3] Faculty of Technology, Department of Electrical Engineering, University Mhamed Bougara of Boumerdes 35000, Algeria.
E-mails: saidi.ahmed@univ-adrar.edu.dz, abdelghani.draoui@cuniv-tindouf.dz, a.touhami@univ-boumerdes.dz

DOI: 10.59277/RRST-EE.2025.70.4. 14

Mathematically, the Fisher criterion seeks to maximize the ratio of the between-class variance to the within-class variance. Specifically, for two classes (*e.g.*, normal and faulty data), the between-class scatter matrix S_b and the within-class scatter matrix S_w are computed as outlined in [34]:

$$S_b = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^{\mathrm{T}}, \tag{1}$$

$$S_w = \sum_{i=1}^{n1} (x_i - \mu_1)(x_i - \mu_1)^{\mathrm{T}} + \sum_{j=1}^{n2} (x_j - \mu_2)(x_j - \mu_2)^{\mathrm{T}}, \tag{2}$$

where μ_1 and μ_2 represent the mean vectors of the two classes, and x_i and x_j are the individual data points within each class. The FRM is used to compute the optimal projection vector that maximizes class separability by maximizing the between-class variance and minimizing the within-class variance.

Between-class variance is a measure of the distinctiveness of the different classes [22]. Within-class variance, conversely, is indicative of the extent to which data points are concentrated within each class.

The objective is to project the data onto a new axis such that the ratio of between-class variance to within-class variance is maximised. This transformation is defined by the Fisher criterion [12].

$$J(w) = \frac{w^{\mathrm{T}} S_b w}{w^{\mathrm{T}} S_w w}. \tag{3}$$

where:

- S_b is the between-class scatter matrix,
- S_w is the within-class scatter matrix,
- w is the projection vector,
- $J(w)$ is the Fisher criterion to be maximized.

2.2 USING THE FISHER RANDOM MATRIX IN FAULT DETECTION

The Fisher random matrix (FRM) improves class separation, a critical advancement in fault detection. By maximising the distinction between healthy and faulty states, FRM improves anomaly detection in PV systems [23]. Key benefits of FRM include

- Dimensionality reduction: PV system data often has high dimensionality (e.g. voltage, current, temperature, irradiance, power). FRM reduces data dimensionality while preserving essential classification information, enabling faster and more efficient fault detection [12].

Robustness to noise: FRM's random matrix elements enhance resilience to noise and data variations common in real PV systems, ensuring accurate fault detection under varying operating conditions [17].

- Scalability: FRM can handle large PV arrays, analysing data from hundreds or thousands of modules simultaneously [20].

After data is transformed using FRM, the Mahalanobis distance classifies new observations as normal or faulty. This distance measures deviations from the healthy class distribution, accounting for variable correlations common in PV monitoring [22].

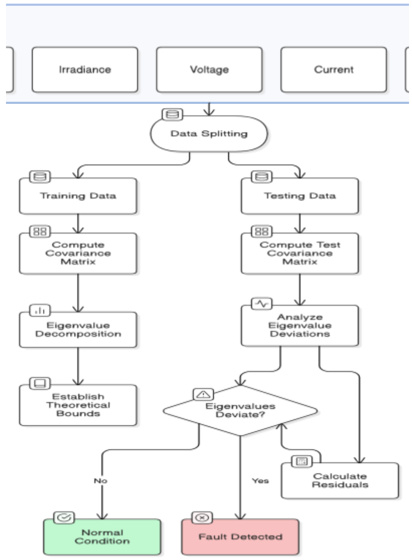

Fig. 1 – Proposed flowchart for fault detection in photovoltaic power systems.

As illustrated by Fig 1, the process of fault detection in photovoltaic (PV) power systems commences with the collection of data, encompassing parameters such as temperature, irradiance, voltage, current, and power. Subsequently, the collected data is segmented into training and testing datasets. The system then computes covariance matrices for both datasets and performs a decomposition of the eigenvalues to analyses the data. Theoretical bounds for the eigenvalues are derived to identify deviations. In the event of detected deviations, residuals are calculated to determine if the system is operating within normal parameters or if a fault has been detected. This process is instrumental in maintaining the reliability and efficiency of power systems by facilitating the swift identification and rectification of any faults.

3. **PV SYSTEM DESCRIPTION**

The validation of the fault detection strategy was conducted through experimental analysis, utilising real-world data from a 20 MWp grid-connected photovoltaic (PV) plant situated in Adrar, Algeria, a region distinguished by its exceptional solar energy potential, with an average daily solar radiation of 5.7 kWh/m^2. The plant, which has been in operation since March 2017, is a large-scale facility comprising 20 PV arrays rated at 1 MW each.

The system's core components are as follows:

- PV Modules: The array consists of 93 subarrays, each containing 44 YINGLI (YL245-29B) modules.
- Inverters: The energy produced by the subarrays is supplied to SUNGROW (5SG 500MX) grid-connected inverters.
- Transformer: A step-up transformer manufactured by SUNTEN is used to connect the inverters to the grid.

The data for the study were collected at 10-minute intervals from January to March 2024 and managed by a supervisory control and data acquisition (SCADA) system. The system under scrutiny was designed to record electrical measurements (DC and AC) from the inverters and meteorological data from dedicated sensors.

The following meteorological sensors were utilised in the study:

- Solar irradiance: The instrument employed was a Kipp &

Zonen CMP21 pyranometer.

- Temperature: The apparatus under consideration is a J-type thermocouple.
- Wind speed: The instrument employed was a WE-100 sonic anemometer.

The fault detection techniques were applied to a PV array that was under observation and comprised 44 YINGLI (YL245P-29B) modules arranged in two parallel strings of 22 modules in series. The electrical parameters of these modules under Standard Test Conditions (STC: 1000 W/m² irradiance, 25°C) were also considered, as shown in Table 1.

Table 1

Electrical parameters of YINGLI (YL245p-29B) PV module at STC

Peak power (W)	Voltage at maximum power (V)	Current at maximum power (A)	Open circuit voltage (V)	Short circuit current (A)
245	29.6	8.28	37.5	8.83

In this study, the data set includes five critical variables essential for monitoring and diagnosing PV system performance: voltage (V), current (A), power (W), temperature (°C) and irradiance (W/m²). Each parameter is essential for assessing system functionality and identifying potential faults. To explore the relationships between these variables, exploratory data analyses were performed, including correlation matrices, variable distributions and box plots. These methods provide insight into data patterns and interactions, aiding anomaly detection and system optimisation.

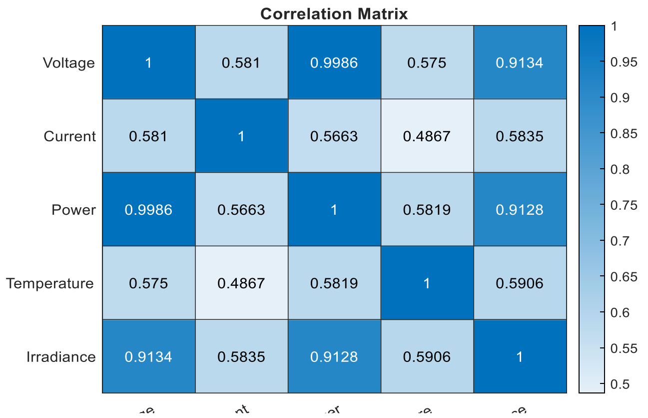

Fig. 2 – Correlation matrix of training and testing data.

The correlation matrix in Fig. 2 shows linear relationships between PV system parameters. Voltage and power show a very strong positive correlation (0.9986), indicating that power output increases significantly with higher voltages. Similarly, voltage and irradiance are highly correlated (0.9134), suggesting that irradiance has a strong effect on voltage output. In contrast, current and temperature have weaker correlations with other parameters, reflecting their more independent behavior under varying conditions.

As illustrated by the box plots in Fig 3, the central tendency, spread, and presence of outliers for each variable are visually represented. For instance, voltage and power demonstrate wide variability, with notable interquartile ranges and a few extreme values. Conversely, temperature has a smaller range of variability, although one extreme outlier is observed below 0°C. These box plots serve to highlight potential outliers, variability in sensor measurements, and deviations that could be indicative of system anomalies or faults.

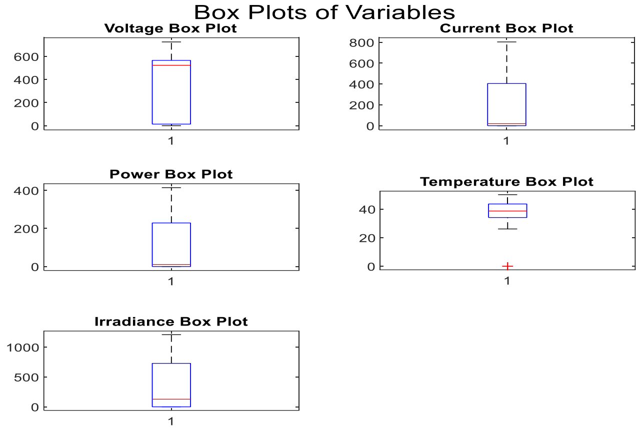

Fig. 3 – Box plots of training and set data.

This descriptive analysis establishes a clear understanding of the data prior to implementing the fault detection algorithms, ensuring that potential trends and irregularities are well accounted for.

4. RESULTS AND DISCUSSION

This section evaluates the performance of the Random Matrix Method (RMM) in detecting DC side faults in a PV system. Six fault types are investigated: 1) string faults, 2) partial shading, 3) module degradation, 4) line-line faults, 5) shorted PV modules, and 6) open circuits in the PV array (see Fig. 4). Experimental data was collected from the PV system described in Section 3.

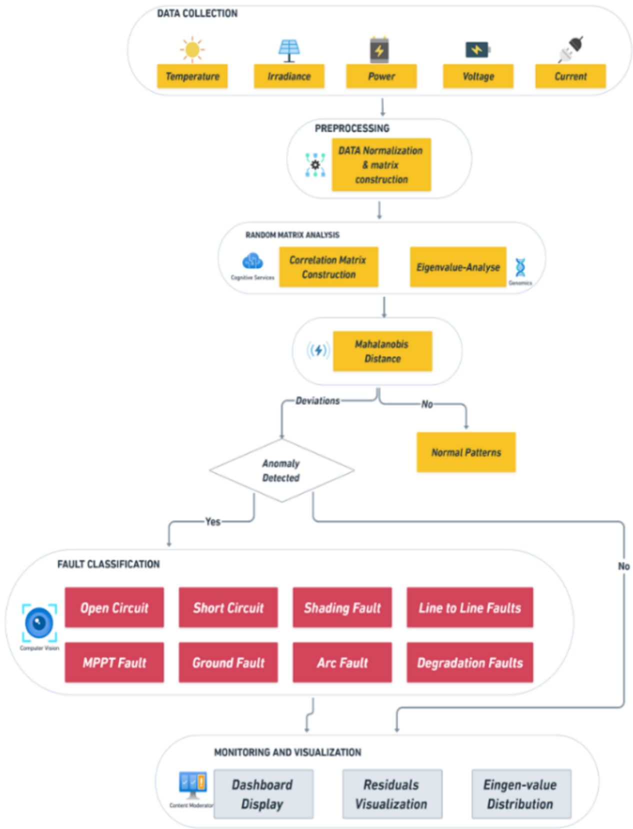

Fig. 4 – RFM-Mahalanobis fault detection and classification proposed in the PV systems flowchart.

The study uses Fisher's linear discriminant (FLD) and Mahalanobis distance methods for fault detection and diagnosis. Figure 4 outlines the RMT-Mahalanobis process, which includes several steps:

1. Data collection: Parameters such as temperature, irradiance, power, voltage and current are collected.
2. Preprocessing: Data is normalised and matrices are constructed.
3. Random matrix analysis: Correlation matrices are constructed, and eigenvalue analysis is performed.
4. Error detection: Mahalanobis distance identifies deviations from normal patterns.
5. Fault classification: Detected anomalies are classified into fault types, including open circuit, short circuit, ground fault, shading fault, line-to-line fault, MPPT fault, ACC fault and degradation fault.
6. Monitoring and visualisation: A dashboard displays residuals, eigenvalue distributions and other metrics for comprehensive system monitoring and fault management.

This process ensures robust fault detection and classification, improving PV system reliability and efficiency. The integration of FLD and Mahalanobis distance methods provides a powerful tool for identifying and managing a wide range of faults, as demonstrated by the experimental results.

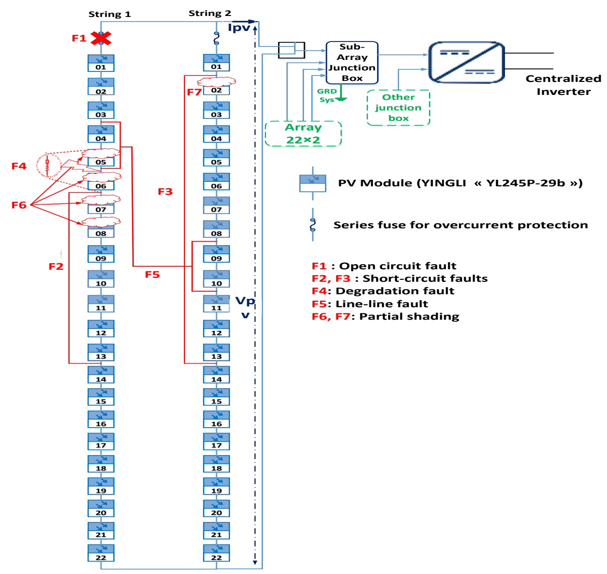

Fig. 5 – Considered faults.

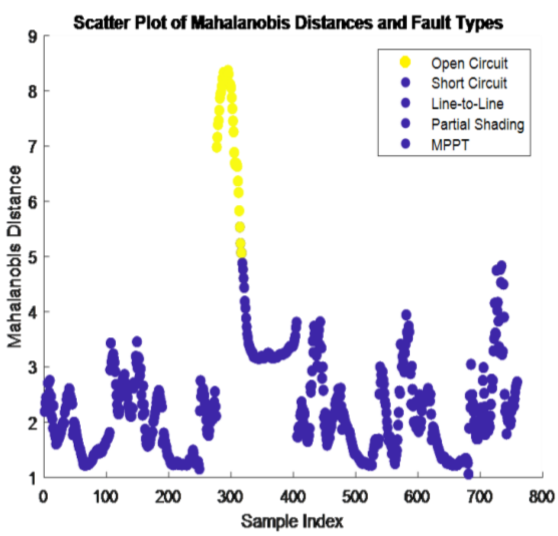

Fig 6 – Scatter plot of Mahalanobis distances and fault type.

This section is devoted to evaluating the performance of RFM to detect faults on the DC side of a PV system. In this study, six types of anomalies are examined: 1) string fault; 2) partial shading; 3) module degradation; 4) line-line faults; 5) PV modules short-circuited; and 6) Open circuit in the PV array in Fig. 5. Experimental data were collected from the PV system described in this section.

The scatter plot in Fig. 6 demonstrates the efficacy of the proposed methodology in multivariate fault classification using the Mahalanobis distance metric. This distance is indicative of the deviation of real-time operational data from the healthy system model, with a higher distance corresponding to a greater anomaly severity. The distribution of data points across the plot demonstrates robust fault discrimination. Specifically, the open circuit faults (yellow points) register the highest distances (peaking above 8), clearly setting them apart from the baseline noise and other fault types (blue points). This visualisation confirms that the method cannot only detect an abnormality but also use the magnitude of the multivariate deviation to accurately isolate and differentiate between distinct system failure modes.

4.1 FISHER'S LINEAR DISCRIMINANT FOR FAULT DETECTION

Fisher's linear discriminant (FLD) is a dimensionality reduction technique that aims to find the optimal hyperplane that maximises the separation between classes while minimising the intra-class variance. It projects high-dimensional data into a lower-dimensional space while preserving class separability. In PV systems, FLD improves fault detection by efficiently distinguishing between healthy and faulty states. By projecting data onto a one-dimensional line, FLD maximises the separation between faulty and healthy classes, thereby improving detection accuracy.

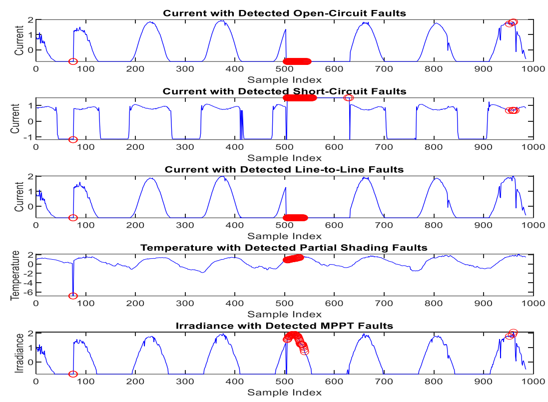

Fig. 7 – Current, temperature, irradiance Time series with detected faults.

A key observation is that Fisher's RM model's performance depends on the separability between faulty and healthy data points. As shown in Fig. 6, applying FLD to PV system data yields clear class separation, thereby enhancing the model's diagnostic capabilities. This highlights the model's reliability in detecting even subtle anomalies.

Mahalanobis distance measures the distance between a point and a distribution, accounting for correlations among variables. It is highly effective at detecting outliers and anomalies in multivariate systems. In PV systems, as shown in Fig. 7, faults are identified by analyzing variations in current, temperature, and irradiance across different fault scenarios, including open-circuit, short-circuit, line-to-line, partial shading, and MPPT faults. Red circles indicate fault locations, and fault parameters are defined by thresholds.

As shown in Fig. 9, the system excels in real-time monitoring, detecting a wide range of fault conditions through detailed analysis of power, temperature, and irradiance data. This allows timely corrective action to prevent further damage or energy loss. The practical utility of the method in maintaining the reliability and efficiency of PV systems underlines its importance for real-time fault detection and system optimization.

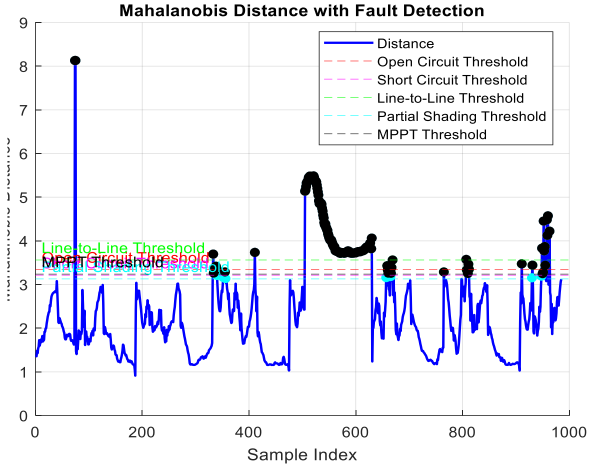

Fig. 8 – Mahalanobis distance threshold categories: fault detected.

The Mahalanobis distance is used in PV systems to detect faults by measuring the multivariate distance of each sample from a healthy reference state, as illustrated in Fig. 8. The distance, measured on the y-axis, is then compared against fault-specific thresholds. Peaks that exceed these thresholds are indicative of defects, with larger peaks indicating severe problems and smaller ones minor anomalies. Under typical operating conditions, the distance fluctuates around a baseline, but deviations caused by faults trigger alerts for investigation.

This method facilitates real-time monitoring by establishing customised thresholds for various faults, thereby ensuring early detection and diagnosis. The efficacy of the method is illustrated in Fig. 9, where multivariate distance successfully differentiates between fault types. The method involves comparing data points to a reference distribution to identify deviations from normal operation. The primary advantages of this system include the capacity to detect minor defects, such as partial shading and degradation, at an early stage. This capability facilitates the implementation of preventive maintenance measures, thereby minimising periods of downtime. The apparatus under scrutiny in this study has been demonstrated to possess the capability of accurately identifying faults of a serious nature, including but not limited to open circuits, short circuits, and arc faults. This capacity is instrumental in preventing potential damage.

The Mahalanobis distance-based approach is robust and reliable, using customised thresholds for accurate fault detection and minimising false positives. The system's adaptability facilitates the implementation of specific monitoring procedures tailored to the characteristics of each system. The efficacy of the system in detecting a wide range of faults has been well documented, including electrical problems (e.g. open and short circuits) and operational problems (e.g. MPPT failures, component degradation). The capacity of this system to enhance system reliability, reduce energy dissipation, and ensure safety is further substantiated by Fig. 8 and 9.

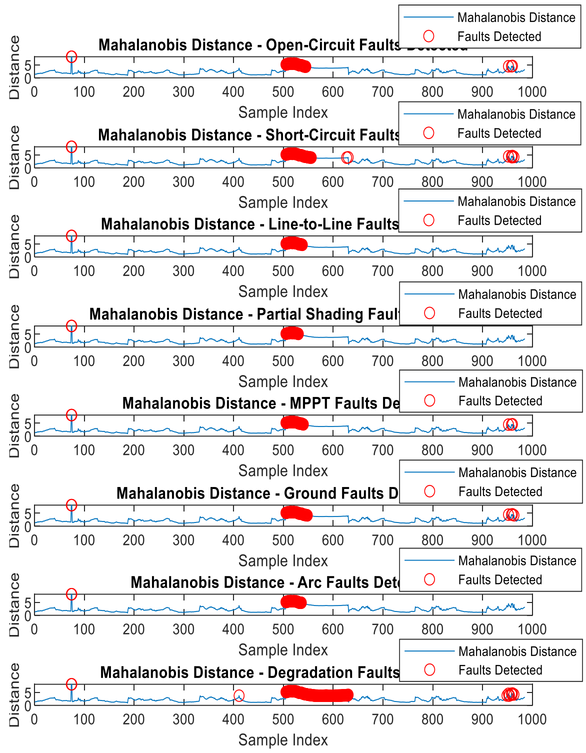

Fig. 9 – Mahalanobis distance vs. multi-classification faults type.

5. CONCLUSIONS AND FUTURE WORKS

The paper introduces and validates a highly effective Fault Detection and Diagnosis strategy for Photovoltaic systems by integrating Fisher's Linear Discriminant with the Mahalanobis Distance Metric (DM).

The proposed methodology has been demonstrated to deliver superior performance by systematically addressing key challenges in PV monitoring.

- Dimensionality and separation: The FLD model has been shown to effectively maximise the separation between healthy operation and fault states, while concomitantly reducing the dimensionality of complex, high-dimensional data. This simplification enhances the model's computational efficiency and clarity.
- Robust diagnosis: The Mahalanobis distance builds on this data structure by accounting for multivariate correlations. This approach facilitated precise classification and a nuanced understanding of diverse fault types, including open circuit, shading, and short circuit failures.

The utilisation of a combined FLD-DM approach results in the generation of a robust and reliable FDD solution. This solution is distinguished by its capacity to manage correlated inputs and preserve remarkably low false negative rates. This integration results in the creation of a powerful, practically applicable tool that is essential for safeguarding the reliability and efficiency of large-scale PV installations.

To build upon the success of the integrated FLD-DM approach, future research will focus on the following key areas:

- Real-time edge implementation: The objective of this project is to develop and test a lightweight version of the FLD-DM model that is suitable for deployment directly onto edge computing devices, such as inverters or local data loggers. This will facilitate expedited decision-making and mitigate the latency associated with cloud-based analytics.

- Adaptive Thresholding and machine learning integration: It is imperative to implement an adaptive mechanism that can dynamically adjust the DM fault thresholds in accordance with real-time environmental conditions, such as irradiance and temperature. Furthermore, the integration of FLD-reduced features into deep learning classifiers should be investigated to enhance the accuracy of distinguishing between subtle fault subtypes and degradation patterns.
- Localization and remediation: The FDD framework is to be extended to include fault localization capabilities (e.g., identifying the specific string or module affected). Future research should also encompass the development of recommended remediation actions for operators to transition the system from a rudimentary diagnostic capacity to a comprehensive, intelligent maintenance platform.

ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to Ahmed Draia University – Adrar for providing the necessary resources and support to conduct this research. Special thanks to the LDDI Laboratory for their technical assistance and collaboration. We extend our thanks to the anonymous reviewers for their constructive comments, which have helped enhance the quality of this manuscript.

CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

Ahmed Saidi: data collection, conceptualization, methodology, writing - Original Draft, Formal Analysis, Software, and validation.
Draoui Abdelghani: data curation, visualization, software, validation.
Abdelouahed Touhami: investigation, writing – review & editing, resources

Received 12 January 2025

REFERENCES

1. B. Gong et al., *Fault diagnosis of photovoltaic array with multi-module fusion under hyperparameter optimization*, Energy Conversion and Management, **319** (2024).
2. B. Ponnuswamy, C. Columbus, S.R. Lakshmi, and J. Chithambaram, *Wind turbine fault modeling and classification using cuckoo-optimized modular neural networks*, Rev. Roum. Sci. Techn. – Électrotechn. et Énerg., **68**, *4*, pp. 369–374 (2023).
3. C. Saiprakash et al., *Improved fault detection and classification in PV arrays using Stockwell transform and data mining techniques*, Results in Engineering, **23** (2024).
4. K. Ding et al., *Feature extraction and fault diagnosis of photovoltaic array based on current–voltage conversion*, Applied Energy, **353** (2024).
5. F. Harrou, A. Saidi, Y. Sun, and S. Khadraoui, *Monitoring of Photovoltaic Systems Using Improved Kernel-Based Learning Schemes*, IEEE Journal of Photovoltaics, **11**, *3*, pp. 806–818 (2021).
6. E.Q.J. M. and V. Subramanian, *Strategic load flow assessment for a ship microgrid with photovoltaic power integration*, Rev. Roum. Sci. Techn. – Électrotechn. et Énerg., **70**, *3*, pp. 397–402 (2025).
7. A. Saidi, B. Cherif, and B. Chellali, *Fuzzy intelligent control for solar/wind hybrid renewable power system*, EEA - Electrotehnica, Electronica, Automatica, **65**, *4* (2017).
8. R. Tang et al., *Fault classification of photovoltaic module infrared images based on transfer learning and interpretable convolutional neural network*, Solar Energy, **276** (2024).
9. Z. Hu et al., *Improved multistep ahead photovoltaic power prediction model based on LSTM and self-attention with weather forecast data*, Applied Energy, **359** (2024).
10. M. Jalal et al., *Deep learning approaches for visual faults diagnosis of photovoltaic systems: State-of-the-Art review*, Results in Engineering, **23** (2024).
11. A. Keddouda et al., *Photovoltaic module temperature prediction using various machine learning algorithms: Performance evaluation*, Applied Energy, **363** (2024).
12. D. Li et al., *Sensing anomaly of photovoltaic systems with sequential conditional variational autoencoder*, Applied Energy, **353** (2024).
13. Q. Qi et al., *Development assessment of regional rooftop photovoltaics based on remote sensing and deep learning*, Applied Energy, **375** (2024).
14. T. Peng et al., *An integrated framework of Bi-directional long-short term memory (BiLSTM) based on sine cosine algorithm for hourly solar radiation forecasting*, Energy, **221** (2021).
15. J. Zhang et al., *Fast object detection of anomaly photovoltaic (PV) cells using deep neural networks*, Applied Energy, **372** (2024).
16. H. Wang et al., *Multi-prediction of electric load and photovoltaic solar power in grid-connected photovoltaic system using state transition method*, Applied Energy, **353** (2024).
17. E.A. Ramadan et al., *An innovative transformer neural network for fault detection and classification for photovoltaic modules*, Energy Conversion and Management, **314** (2024).
18. M.F. Tahir et al., *Enhancing PV power forecasting with deep learning and optimizing solar PV project performance with economic viability: A multi-case analysis of 10 MW Masdar project in UAE*, Energy Conversion and Management, **311** (2024).
19. A. Słowik et al., *An efficient approach to parameter extraction of photovoltaic cell models using a new population-based algorithm*, Applied Energy, **364** (2024).
20. L.-A. Mustaţă, G.-D. Sorea, and E. Helerea, *Monitoring and control software for a standalone house supply system*, Rev. Roum. Sci. Techn. – Électrotechn. et Énerg., **70**, *2*, pp. 169–174 (2025).
21. E.Q.J. M. and V. Subramanian, *Strategic load flow assessment for a ship microgrid with photovoltaic power integration*, Rev. Roum. Sci. Techn. – Électrotechn. et Énerg., **70**, *3*, pp. 397–402 (2025).
22. L. Khettache, M.R. Rezoug, A. Saadi, and L. Sahraoui, *A novel design of a photovoltaic system based on a linear induction motor and a reciprocating pump*, Rev. Roum. Sci. Techn. – Électrotechn. et Énerg., **70**, *1*, pp. 3–8 (2025).