DATA-DRIVEN FAULT DETECTION IN PHOTOVOLTAIC SYSTEMS USING MACHINE LEARNING

Nameera Jani[1], Dr. Mohammed Pasha Professor[2]
[1]PG Scholar, Department of CS, Shadan Women's College of Engineering and Technology, Hyderabad, nameerajani1210@gmail.com
[2]Professor, Department of CSE, Shadan Women's College of Engineering and Technology
mohd.pasha@outlook.com

ABSTRACT:
For sustainable energy production, photovoltaic (PV) systems must operate dependably and efficiently. However, flaws like dirt buildup and partial shade drastically lower PV modules' power production. This research offers a machine learning-based fault detection framework that uses just electrical and environmental data to classify and identify defects in PV systems in order to address this problem.
Logistic regression, decision trees, random forests, support vector machines (SVM), and artificial neural networks (ANN) are among the supervised machine learning methods used in the suggested system.
Important characteristics including voltage, current, ambient temperature, and irradiance are included in the dataset, which was gathered under three different operating conditions: normal operation, partial shading, and dirt accumulation. To get the dataset ready for training and testing, data preparation methods such label encoding, data splitting, and normalisation were used. Standard classification criteria, including accuracy, precision, recall, F1-score, and confusion matrix, were used to assess each model's performance. The Artificial Neural Network achieved the best accuracy, surpassing 98% precision, according to experimental data, which show that all models performed well in identifying errors inside the same system used for training. However, models' performance deteriorated when tested on data from a separate PV system with different features, underscoring the necessity of system-specific model training. This study finds that when trained on pertinent system data, machine learning models—in particular, artificial neural networks—are useful for failure identification in photovoltaic systems. The method provides an automated, economical way to improve solar energy systems' performance and dependability.

1. INTRODUCTION:

Photovoltaic (PV) systems are now an essential part of contemporary power generation due to the growing demand for clean, sustainable energy sources worldwide. The efficiency and dependability of PV systems become essential as the globe moves toward renewable energy to fight environmental deterioration and lessen dependency on fossil fuels. Although these systems are often low-maintenance and convert sunlight into power, their effectiveness is extremely sensitive to external factors including physical obstacles and environmental conditions.
Even though PV systems have advanced technologically, flaws like partial shade, dirt buildup, or hardware problems frequently cause performance decrease in real-world operations. In large-scale installations, these abnormalities frequently go undetected and result in a significant drop in power production, which lowers overall system efficiency and causes energy losses. Conventional defect detection techniques are less practical for scaling deployments because they are either expensive, necessitate manual inspection, or lack the accuracy required for prompt response.
This paper suggests a machine learning-based fault detection framework that uses solely electrical and environmental data to accurately identify faults in order to overcome these limitations. The system divides operating conditions into normal, partially shaded, and dirt-affected categories using supervised learning methods such Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), and Artificial Neural Networks (ANN). This data-driven strategy improves system reliability and guarantees affordable, real-time issue detection. The study emphasises the necessity of system-specific training to sustain high performance as it investigates the generalisability of these models across various PV systems.

1.1 OBJECTIVE:

This project's main goal is to provide a dependable and affordable machine learning-based framework that uses solely electrical and environmental data to identify and categorise defects in photovoltaic (PV) systems. By precisely identifying typical problems like partial shade and dirt collection, which have a substantial impact on energy output, the goal is to improve the operational efficiency and dependability of PV installations. The research aims to identify the best model for fault classification by utilising a variety of supervised machine learning methods, such as Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), and Artificial Neural Network (ANN). The goal also includes studying these models' generalisability across various PV systems and assessing their performance using common categorisation measures. By using this strategy, the project hopes to promote the more general objective of sustainable energy generation by helping to develop sophisticated, automated solar energy monitoring systems.

2. EXISTING SYSTEM:

Fault detection in conventional photovoltaic (PV) systems is frequently carried out by hand inspections or hardware-based monitoring tools like current sensors and thermal cameras. In addition to being labour-intensive and time-consuming, these techniques necessitate substantial maintenance and operating expenses, particularly in large-scale solar farms. Early-stage shading or light dirt deposition are examples of subtle or partial flaws that can progressively impair system performance without setting off alarms, yet conventional procedures frequently miss them. Furthermore, the efficiency of these systems in determining the underlying cause of performance losses is often limited due to their lack of intelligent fault classification capabilities.

Recent developments in data-driven technologies have made it possible to monitor electrical outputs and spot anomalies using statistical methods or threshold-based rule engines. Nevertheless, these techniques have poor accuracy and significant false-positive rates due to their sensitivity to environmental changes like temperature and light. Additionally, they are frequently designed for certain system configurations, which prevents them from being scalable across other PV setups. An intelligent, adaptable, and scalable solution that can precisely identify and categorise errors in real-time without requiring expensive hardware installations is still desperately needed, despite considerable improvements.

2.1 Existing System Disadvantages:

- Prone to overfitting
- Sensitivity to noisy data
- Bias toward features with more levels
- Limited ability to capture complex patterns

3. LITERATURE SURVAY:

Artificial Intelligence in Photovoltaic Fault Identification and Diagnosis: A Systematic Review Author: Mahmudul Islam et al. 2023

This systematic review analyzes 31 AI-based studies for PV fault detection, including ML and DL methods like K-Nearest Neighbor, Random Forest, CNN, and RNN. It finds that deep learning approaches generally outperform traditional ML techniques in terms of accuracy and efficiency, and proposes an ANN-based classification method for PV faults

Development of a machine-learning-based method for early fault detection in photovoltaic systems Author: Stylianos Voutsinas, Dimitrios Karolidis, Ioannis Voyiatzis et al. 2023

This paper presents a logistic regression model (with cross-validation) trained on a year’s worth of irradiance, temperature, voltage, and current data from smart PV arrays. The approach achieves 97.11% accuracy for detecting DC-side faults (e.g., open-circuit, short-circuit, mismatch faults), noting its low computational cost and practicality.

4. RELATED WORK:

Machine learning approaches have been used by a number of academics to identify and categorise faults in photovoltaic (PV) systems. Electrical and environmental characteristics, including as irradiance, ambient temperature, voltage, and current, have been used in a variety of algorithms to detect defects such partial shading (PS), open-circuit (OC), and short-circuit (SC) problems. These research have shown how ML-based methods can improve PV system maintenance and dependability.

Many current studies concentrate mostly on identifying electrical problems, but they fall short in addressing problems brought on by dirt buildup on solar panels. Shading and dirt-related problems are combined into a single category in certain research, which could lower classification accuracy and result in ineffective maintenance choices. In order to provide a more thorough understanding of PV system failures, other researchers have expanded their analysis to include numerous fault circumstances, such as combinations of partial shading with open-circuit and short-circuit faults.

Computer vision techniques have also been used in recent developments to diagnose photovoltaic faults. Defects using image-based data have been identified using Convolutional Neural Networks (CNNs), k-Nearest Neighbours (kNN) algorithms, Neuro-Fuzzy systems, and semantic image segmentation techniques. These methods have demonstrated encouraging outcomes in identifying panel anomalies and enhancing fault classification efficiency. Furthermore, research on concentrated solar power systems has investigated machine learning-based structural fault identification, providing insightful information for comparable applications in photovoltaic systems.

Despite tremendous advancements, the performance and applicability of current methods vary greatly. Many studies' capacity to generalise across various PV systems and operating situations is limited because they depend on simulated datasets or concentrate on particular fault types. Additionally, the ability of models to adapt to real-world situations and the identification of several simultaneous errors continue to be difficult problems. These drawbacks emphasise the necessity of strong machine learning frameworks that can reliably identify and categorise dirt accumulation and partial shading defects using actual sensor data.

5. PROPOSED SYSTEM:

Using electrical and environmental characteristics, the suggested system presents a machine learning-based framework for automatic defect identification in photovoltaic (PV) systems. This system uses data like voltage, current, irradiance, and ambient temperature to identify and categorise issues such partial shading and dirt accumulation, in contrast to conventional approaches that depend on hardware-based monitoring or manual examination. The goal is to provide a low-cost, scalable, and intelligent solution that can lower maintenance costs while increasing energy production efficiency and real-time defect detection.

Several machine learning algorithms, such as Random Forest, Support Vector Machine, Decision Tree, K-Nearest Neighbours, and Logistic Regression, are applied and compared using this system. The dataset is used to train and assess these models following extensive data preprocessing, including cleaning, normalisation, and encoding. Metrics including accuracy, precision, recall, and F1-score are used to gauge performance. Random Forest is suggested as the best option among the assessed models because of its high classification accuracy (more than 98% in testing), resilience to overfitting, and capacity to manage non-linear trends in PV system behaviour. The intelligent defect monitoring system, which can be expanded for use in different PV installations, will be built around this paradigm.

5.1 **Proposed System Advantages**:

- Reduces Over fitting
- Works Well with Missing Data
- Handles Non-Linearity
- Resistant to Noise

6. MODULES:

MODULE EXPLANATION:

Data Collection Module:

The framework for fault detection is built on the Data Collection Module. It entails gathering pertinent data from photovoltaic (PV) systems with an emphasis on environmental and electrical aspects. Voltage, current, irradiance, and ambient temperature are important characteristics that are crucial indications of PV system performance. Three distinct operational scenarios—normal circumstances, partial shading, and dirt accumulation—are used to collect the data. This module may rely on publicly accessible or simulated datasets, or it may use sensors in real-world settings. In order for the models trained in subsequent stages to successfully learn the differences between different fault types, accurate and representative data collection is essential.

Data Preprocessing Module:

After being gathered, the raw data is preprocessed to make sure it is consistent, clean, and appropriate for machine learning model training. This module carries out crucial tasks such encoding categorical labels into machine-readable formats, addressing missing or noisy values, and normalising numerical features to a common scale. Additionally, the dataset is divided into training and testing sets, usually in a conventional ratio like 80:20. By lowering bias and variation in the data and maintaining the integrity of the machine learning pipeline, proper preprocessing enhances model accuracy and generalisation.

Model Training Module:

The preprocessed dataset is used to train a variety of supervised machine learning algorithms in the Model Training Module. Logistic regression, decision trees, random forests, support vector machines (SVM), and artificial neural networks (ANN) are among the models that were chosen. Every algorithm discovers the connections and patterns between the target labels (fault types) and the input features. In order to reduce classification mistakes, internal parameters are optimised during the training phase. This module is essential for creating reliable models that can precisely and automatically identify various PV system conditions.

Model Evaluation Module:

Following training, the models are assessed to see how well they identify errors. The Model Evaluation Module evaluates each algorithm's performance using performance metrics like accuracy, precision, recall, F1-score, and confusion matrix. These measurements shed light on each model's advantages and disadvantages, especially with regard to identifying particular problem kinds without incorrectly classifying them. The best-performing model (such as an ANN) can be chosen for deployment in the fault detection system by comparing models using a standardised evaluation framework.

Fault Detection and Classification Module:

This module is the core component of the operational phase, when incoming data from a PV system is classified using the trained machine learning model. It analyses test or real-time data to ascertain whether the system is functioning regularly or is impacted by dirt buildup or partial shading. This module guarantees accurate defect detection by utilising the model with the best evaluation performance. It makes it possible to quickly and automatically identify problems, minimising downtime and preserving PV installations' energy production efficiency.

Result Visualization:

The model performance and fault classification results are displayed graphically and tabularly in the Result Visualisation Module. Stakeholders are assisted in interpreting the efficacy of the system by tools such as prediction summaries, bar charts of performance measures, and confusion matrices. By improving the user experience and offering transparency into model behaviour, this module makes it simpler to identify possible problems, comprehend system dependability, and make data-driven choices regarding PV maintenance and monitoring tactics.

7. TECHNIQUE USED:

7.1 **EXISTING TECHNIQUE**:

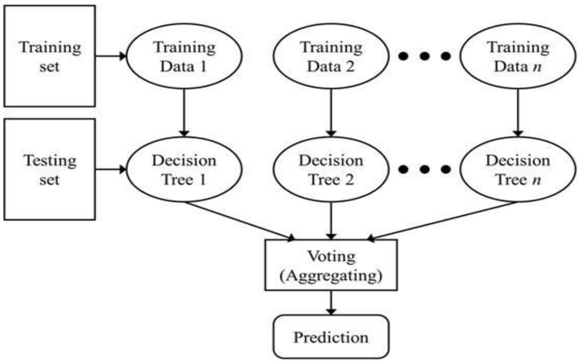
Manual inspections and hardware-based monitoring tools like heat cameras, current sensors, and threshold-based systems are the primary methods now used for problem identification in photovoltaic (PV) systems. By looking for abnormalities in electrical outputs like voltage and current, these techniques usually concentrate on finding obvious or significant defects. However, hardware sensors need costly installation and upkeep, while manual examinations are time-consuming, expensive, and prone to human error.

Furthermore, a lot of current techniques rely on statistical analysis or basic threshold criteria, which are frequently vulnerable to environmental variations like temperature swings and fluctuating irradiance, resulting in overlooked defects or false alarms.
Additionally, these conventional methods are limited in their use for preventive maintenance since they are unable to precisely define fault types, such as differentiating between partial shading and dirt accumulation. Additionally, a lot of methods are made for particular PV system configurations and do not adapt well to other module features or environmental circumstances. Because of this, they provide little flexibility and scalability in actual, diverse PV installations.

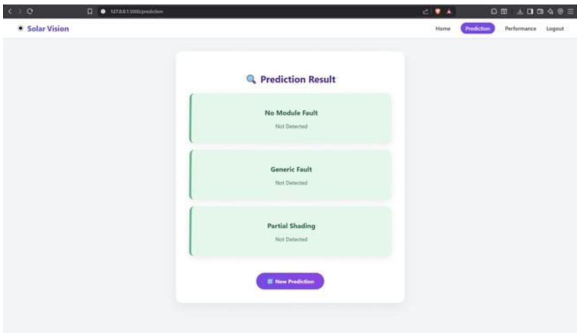
7.2 PROPOSED TECHNIQUE:

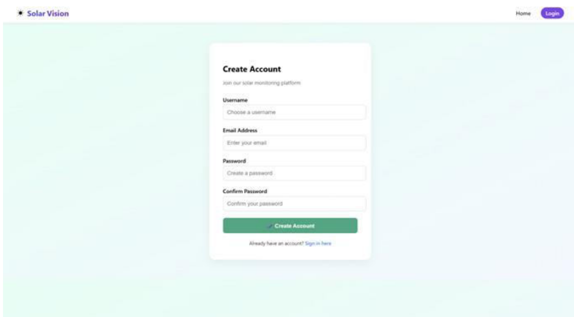
By applying machine learning (ML) techniques to identify and categorise PV system defects using solely electrical and environmental sensor data, the suggested solution overcomes the drawbacks of current approaches. The system learns intricate patterns in data that describe both normal operation and fault states like partial shading and dirt accumulation by utilising supervised learning models including Random Forest, Support Vector Machine, and Logistic Regression. This data-driven method enables automatic,
real-time defect identification with greater accuracy and robustness by reducing reliance on expensive hardware and manual examination.
Additionally, the suggested approach is made to be scalable and adaptive. The system can better generalise across various PV setups by employing features like voltage, current, irradiance, and ambient temperature and training models on a variety of datasets. Machine learning models, particularly Random Forest, can achieve precision values exceeding 98%, greatly improving fault classification and enabling proactive maintenance strategies that improve PV system reliability and energy output, according to performance evaluation using metrics like accuracy, precision, and recall.

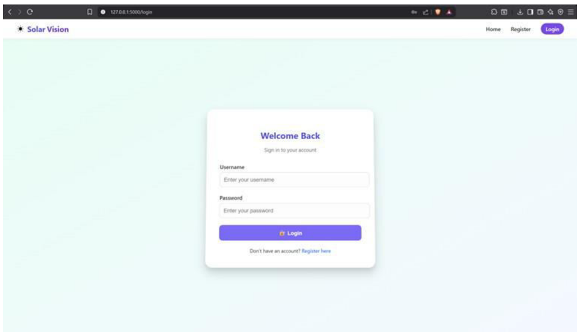
7.3 SYSTEM ARCHITECTURE:

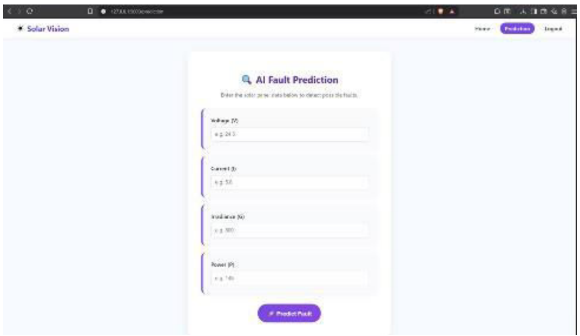


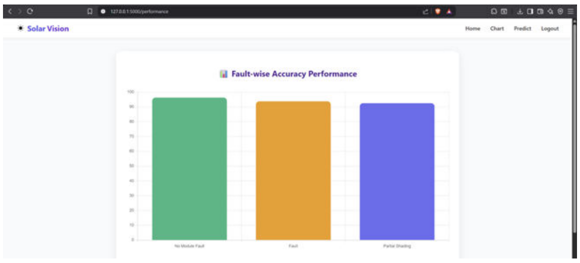
8. RESULT

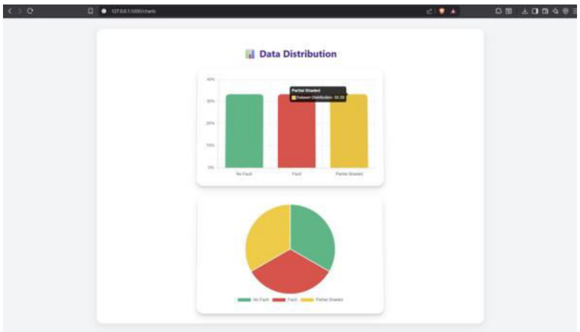












9. DISCUSSION AND CONCLUSSION:

Using solely electrical and environmental data, this experiment shows how machine learning, namely the Random Forest method, may be effectively used to identify and categorise defects in photovoltaic (PV) systems. The system effectively detects typical problems like partial shade and dirt accumulation, which are known to drastically lower solar energy output, by concentrating on important factors like voltage, current, temperature, and irradiance. The model's accuracy and robustness under the identical system conditions used for training are guaranteed by the use of supervised learning and preprocessing procedures.

According to experimental findings, the Random Forest model offers a scalable, automated, and economical way to improve the performance and dependability of PV systems by classifying faults with high accuracy. However, because model performance can deteriorate when tested on data from various PV systems, the experiment also emphasises the necessity of system-specific training. All things considered, the study supports the promise of machine learning in intelligent solar energy management, opening the door for proactive maintenance and intelligent defect identification in actual PV installations.

10. FUTURE ENHANCEMENTS:

Developing a real-time monitoring system coupled with Internet of Things (IoT) devices for real-time data collecting from PV systems would be a significant improvement in further versions of this project. The system can quickly identify anomalies and errors by using wireless sensors to continuously monitor variables including voltage, current, temperature, and irradiance. This reduces downtime and boosts operational efficiency. Furthermore, especially in remote solar farms, integrating edge computing for initial processing at the data source might lower latency and enable prompt failure reactions.

Enhancing the model's ability to generalise by training it on a variety of datasets from various hardware and geographic situations is another significant improvement that will enable it to adapt to different PV system kinds and configurations. Convolutional neural networks (CNNs) and recurrent neural networks (RNNs), two deep learning techniques, may be used to more successfully capture intricate patterns and time-series behaviours. Additionally, the system would be a complete solution for PV system operators if it were integrated into a web or mobile dashboard with performance metrics and predictive maintenance notifications.

REFERENCES:

[1] A. Kumaradurai, Y. Teekaraman, T. Coosemans, and M. Messagie, "Fault.detection in photovoltaic systems using machine learning algorithms: Areview," in Proc. 8th Int. Conf. Orange Technol. (ICOT), Dec. 2020,pp. 1–5.

[2] M. Meribout, "Sensor systems for solar plant monitoring," IEEE Trans.Instrum. Meas., vol. 72, pp. 1–16, 2023.

[3] Y. Higuchi and T. Babasaki, "Failure detection of solar panels usingthermographic images captured by drone," in Proc. 7th Int. Conf. Renew.Energy Res. Appl. (ICRERA), Oct. 2018, pp. 391–396.

[4] Y. Zhao, L. Yang, B. Lehman, J.-F. de Palma, J. Mosesian, and R. Lyons,"Decision tree-based fault detection and classification in solar photovoltaicarrays," in Proc. Twenty-Seventh Annu. IEEE Appl. Power Electron. Conf.Expo. (APEC), Feb. 2012, pp. 93–99.

[5] S. U. Ahmed, M. Affan, M. I. Raza, and M. Harris Hashmi, "Inspectingmega solar plants through computer vision and drone technologies," in Proc. Int. Conf. Frontiers Inf. Technol. (FIT), Dec. 2022, pp. 18–23.41420 VOLUME 13, 2025J. Zwirtes et al.: Fault Detection in Photovoltaic Systems Using a Machine Learning Approach

[6] C. Henry, S. Poudel, S.-W. Lee, and H. Jeong, "Automatic detection system of deteriorated PV modules using drone with thermal camera," Appl. Sci.,vol. 10, no. 11, p. 3802, May 2020.

[7] L. Yun, Y. Bofeng, Q. Dan, and L. Fengshuo, "Research on fault diagnosis of photovoltaic array based on random forest algorithm," in Proc. IEEE Int. Conf. Power Electron., Comput. Appl. (ICPECA), Jan. 2021, pp. 194–198.

[8] M. Dhimish and G. Badran, "Photovoltaic hot-spots fault detection algorithm using fuzzy systems," IEEE Trans. Device Mater. Rel., vol. 19, no. 4, pp. 671–679, Dec. 2019.

[9] A. Kumari, A. Shekhar, and M. S. Kumar, "An artificial neural networkbased fault detection technique for PV array," in Proc. 2nd Int. Conf.Emerg. Frontiers Electr. Electron. Technol. (ICEFEET), Jun. 2022,pp. 1–6.

[10] F. Aziz, A. Ul Haq, S. Ahmad, Y. Mahmoud, M. Jalal, and U. Ali, "Anovel convolutional neural network-based approach for fault classificationin photovoltaic arrays," IEEE Access, vol. 8, pp. 41889–41904, 2020.

[11] W. Gao, "PV array fault detection based on deep neural network," in Proc.IEEE Green Technol. Conf. (GreenTech), Apr. 2021, pp. 42–47.

[12] D. Wang, Y. Wang, W. Li, and H. Dong, "Distributed photovoltaic power station fault diagnosis based on random forest," in Proc. 9th Int. Conf.Control, Mechatronics Autom. (ICCMA), Nov. 2021, pp. 90–94.

[13] N.-C. Yang and H. Ismail, "Robust intelligent learning algorithm usingrandom forest and modified-independent component analysis for PVfault detection: In case of imbalanced data," IEEE Access, vol. 10,pp. 41119–41130, 2022.

[14] Z. Chen, Y. Chen, L. Wu, S. Cheng, and P. Lin, "Deep residual network based fault detection and diagnosis of photovoltaic arrays using currentvoltage curves and ambient conditions," Energy Convers. Manage.,vol. 198, Oct. 2019, Art. no. 111793.

[15] M. M. Mansouri, S. Hadjeri, and M. Brahami, "New method of detection, identification, and elimination of photovoltaic system faults in real time based on the adaptive neuro-fuzzy system," IEEE J. Photovolt., vol. 11,no. 3, pp. 797–805, May 2021.

[16] Z. Yi and A. H. Etemadi, "Fault detection for photovoltaic systems based on multi-resolution signal decomposition and fuzzy inference systems,"IEEE Trans. Smart Grid, vol. 8, no. 3, pp. 1274–1283, May 2017.

[17] J. Barbosa, R. L. P. de Medeiros, F. A. C. Ayres, J. E. C. Filho, V. F. Lucena,and I. Bessa, "Fault detection for photovoltaic systems using fuzzy Cmeans clustering," in Proc. IEEE 27th Int. Conf. Emerg. Technol. Factory Autom. (ETFA), Sep. 2022, pp. 1–5.

[18] S. Fadhel, D. Diallo, C. Delpha, A. Migan, I. Bahri, M. Trabelsi, and M. F. Mimouni, "Maximum power point analysis for partial shading detection and identification in photovoltaic systems," Energy Convers. Manage., vol. 224, Nov. 2020, Art. no. 113374.

[19] W. Al Abri, R. A. Abri, H. Yousef, and A. Al-Hinai, "A simple method for detecting partial shading in PV systems," Energies, vol. 14, no. 16,p. 4938, Aug. 2021.

[20] Y. Higuchi and T. Babasaki, "Classification of causes of broken solar panels in solar power plant," in

Proc. IEEE Int. Telecommun. Energy Conf. (INTELEC), Oct. 2017, pp. 127–132.

[21] S. Leva, M. Aghaei, and F. Grimaccia, "PV power plant inspection by UAS: Correlation between altitude and detection of defects on PV modules," in Proc. IEEE 15th Int. Conf. Environ. Electr. Eng. (EEEIC), Jun. 2015, pp. 1921–1926.

[22] K. A. Abuqaaud and A. Ferrah, "A novel technique for detecting and monitoring dust and soil on solar photovoltaic panel," in Proc. Adv. Sci. Eng. Technol. Int. Conf. (ASET), Feb. 2020, pp. 1–6.

[23] F. E. Alfaris, "A sensorless intelligent system to detect dust on PV panels for optimized cleaning units," Energies, vol. 16, no. 3, p. 1287, Jan. 2023.

[24] J. Pinho and M. Galdino, "Manual de engenharia para sistemas fotovoltaicos–cresesb," 2nd edition, São Paulo, Brazil, Tech. Rep. CEPELDTE-CRESESB, 2014.

[25] S. Laamami, M. Benhamed, and L. Sbita, "Artificial neural network-based fault detection and classification for photovoltaic system," in Proc. Int. Conf. Green Energy Convers. Syst. (GECS), Mar. 2017,
pp. 1–7.

[26] S. Gong, X. Wu, and Z. Zhang, "Fault diagnosis method of photovoltaic array based on random forest algorithm," in Proc. 39th Chin. Control Conf. (CCC), Jul. 2020, pp. 4249–4254.

[27] A. Belaout, F. Krim, and A. Mellit, "Neuro-fuzzy classifier for fault detection and classification in photovoltaic module," in Proc. 8th Int. Conf. Model., Identificat. Control (ICMIC), Nov. 2016, pp. 144–149.

[28] R. Cavieres, R. Barraza, D. Estay, J. Bilbao, and P. Valdivia-Lefort, "Automatic soiling and partial shading assessment on PV modules through RGB images analysis," Appl. Energy, vol. 306, Jan. 2022, Art. no. 117964.

[29] M. S. H. Onim, Z. M. M. Sakif, A. Ahnaf, A. Kabir, A. K. Azad, A. M. T. Oo, R. Afreen, S. T. Hridy, M. Hossain, T. Jabid, and M. S. Ali, "SolNet: A convolutional neural network for detecting dust on solar panels," Energies, vol. 16, no. 1, p. 155, Dec. 2022.

[30] T. Cruz-Rojas, J. A. Franco, Q. Hernandez-Escobedo, D. Ruiz-Robles, and J. M. Juarez-Lopez, "A novel comparison of image semantic segmentation techniques for detecting dust in photovoltaic panels using machine learning and deep learning," Renew. Energy, vol. 217, Nov. 2023, Art. no. 119126.

[31] L. Bonsignore, M. Davarifar, A. Rabhi, G. M. Tina, and A. Elhajjaji, "Neuro-fuzzy fault detection method for photovoltaic systems," Energy Proc., vol. 62, pp. 431–441, Jan. 2014.

[32] F. Harrou, B. Taghezouit, and Y. Sun, "Improved kNN-based monitoring schemes for detecting faults in PV systems," IEEE J. Photovolt., vol. 9, no. 3, pp. 811–821, May 2019.