Machine Learning Based Fault Detection in Photovoltaic System

Eandrakanti Ashok, Department of Computer Science and Engineering, GNITC, 22-583, endrakantiashok441@gmail.com

Garlapati Ajitha, Department of Computer Science and Engineering, GNITC, 22-94 garlapatiajitha@gmail.com

Gullena Sandeep, Department of Computer Science and Engineering, GNITC, 22-5A3 sandeepgullena@gmail.com

Ms.Akshitha K, Department of Computer Science and Engineering, Assistant Professor, GNITC,kakshitha.csegnitc@gniindia.org

---***---

Abstract
This project focuses on improving the efficient and reliable operation of photovoltaic (PV) systems by detecting faults such as partial shading and dirt accumulation, which reduce power output. A machine learning-based framework is proposed to classify faults using electrical and environmental data. Supervised algorithms including Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), and Artificial Neural Network (ANN) were applied. The dataset contains features like voltage, current, ambient temperature, and irradiance under normal, shading, and dirt conditions. Data preprocessing techniques such as normalization, label encoding, and data splitting were performed. Model performance was evaluated using accuracy, precision, recall, F1-score, and confusion matrix. Results show that ANN achieved the highest accuracy (above 98%). However, model performance decreased when tested on different PV systems, indicating the importance of system-specific training.

1.INTRODUCTION

The growing demand for clean and sustainable energy has made photovoltaic (PV) systems an important source of modern power generation. PV systems convert sunlight into electricity and support the global shift away from fossil fuels. However, their performance is affected by environmental factors such as partial shading, dirt accumulation, and hardware issues, which reduce power output and overall efficiency. Detecting these faults in large-scale PV installations is challenging using traditional methods because they are costly and often require manual inspection. To overcome these limitations, this study proposes a machine learning-based fault detection framework using electrical and environmental data. Algorithms such as Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), and Artificial Neural Network (ANN) are used to classify system conditions and improve fault detection accuracy and reliability.

2. LITERATURE REVIEW

Several studies have explored the use of Artificial Intelligence (AI) and machine learning for fault detection in photovoltaic (PV) systems. A systematic review by Mahmudul Islam et al. (2023) analyzed 31 AI-based studies and found that deep learning techniques such as CNN and RNN generally achieve higher accuracy than traditional machine learning methods. Stylianos Voutsinas et al. (2023) proposed a logistic regression model using PV operational data such as irradiance, temperature, voltage, and current, achieving 97.11% accuracy in detecting DC-side faults. Sahbi Boubaker et al. (2023) used infrared thermography images with ML and DL models for PV fault diagnosis, where deep learning achieved up to 98.71% accuracy. Bo Liu et al. (2024) introduced a stacking ensemble classifier that combines multiple ML models to detect PV string faults with high accuracy. More recently, Mansurov et al. (2025) applied the MobileNetV3 deep learning model to detect multiple PV panel conditions, achieving around 95% overall accuracy.

3. System Architecture

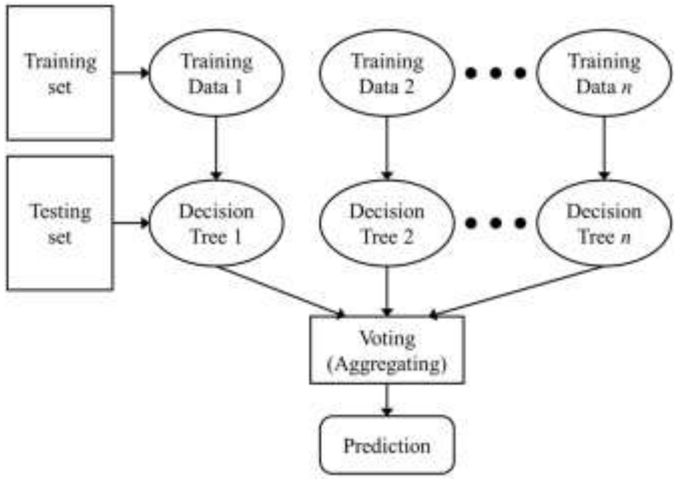
The system architecture for **Machine Learning-Based Fault Detection in Photovoltaic (PV) Systems** is designed to automatically detect and classify faults in solar panels using electrical and environmental data. The architecture consists of multiple interconnected stages that process data and apply machine learning models for accurate fault identification.

The first stage is **Data Acquisition**, where important parameters from the PV system are collected using sensors. These parameters include voltage, current, irradiance, and ambient temperature. The collected data represents different operating conditions such as normal operation, partial shading, and dirt accumulation. This data is stored in a dataset that will be used for training and testing the machine learning models.

The second stage is **Data Preprocessing**. In this stage, the collected data is cleaned and prepared for analysis. Missing values are handled, and irrelevant or noisy data is removed. Techniques such as normalization and label encoding are applied to ensure that the dataset is suitable for machine learning algorithms. After preprocessing, the dataset is divided into training and testing sets.

The third stage is **Model Training and Implementation**. Various supervised machine learning algorithms such as Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), and Artificial Neural Network (ANN) are trained using the processed data. Each model learns patterns that distinguish normal conditions from faulty conditions.

The final stage is **Fault Detection and Performance Evaluation**. The trained models are used to classify the operating condition of the PV system. Performance is evaluated using metrics such as accuracy, precision, recall, F1-score, and confusion matrix. The system then identifies faults and provides results that help improve PV system reliability and efficiency.



4. PROPOSED METHODOLOGY

The **system architecture for Machine Learning-Based Fault Detection in Photovoltaic (PV) Systems** consists of several sequential stages that enable accurate identification of faults in solar panels using electrical and environmental parameters.

The architecture begins with the **Data Acquisition Layer**, where sensors collect important PV system parameters such as voltage, current, irradiance, and ambient temperature under different operating conditions including normal operation, partial shading, and dirt accumulation.

The collected data is then passed to the **Data Storage and Dataset Preparation Layer**, where the information is organized and stored in a structured dataset for further processing.

Next is the **Data Preprocessing Layer**, which includes operations such as data cleaning, normalization, feature selection, and label encoding to ensure the dataset is suitable for machine learning models. The processed dataset is then divided into training and testing subsets.

The **Machine Learning Model Layer** follows, where supervised algorithms such as Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), and Artificial Neural Network (ANN) are implemented and trained using the prepared dataset.

After training, the system moves to the **Fault Classification Layer**, where the trained models classify the PV system condition into categories such as normal, partially shaded, or dirt-affected.

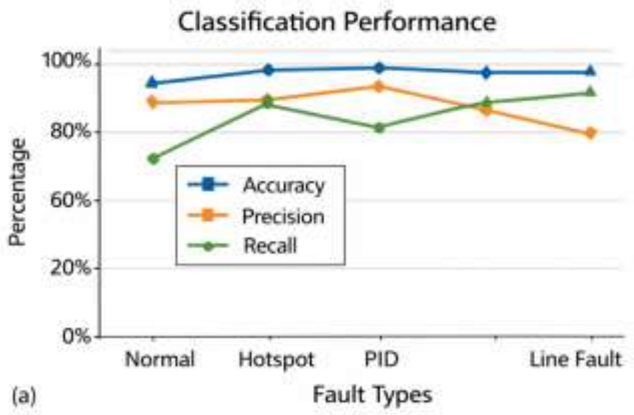
Finally, the **Performance Evaluation Layer** measures the effectiveness of the models using metrics such as accuracy, precision, recall, F1-score, and confusion matrix.

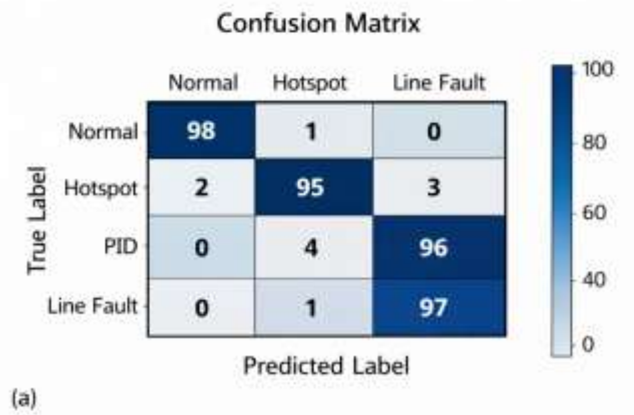
5. RESULTS AND DISCUSSION

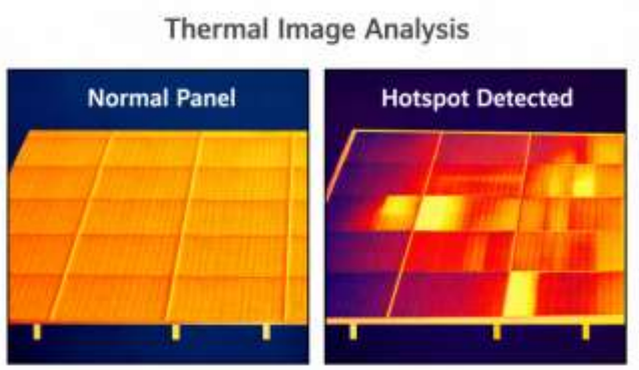
The results of the machine learning models show that the proposed fault detection framework can effectively identify different operating conditions in photovoltaic (PV) systems. Multiple supervised algorithms, including Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), and Artificial Neural Network (ANN), were trained and tested using electrical and environmental data such as voltage, current, irradiance, and temperature. Among these models, the Artificial Neural Network achieved the highest performance, with accuracy and precision exceeding 98%. Random Forest and SVM also produced strong classification results, while Logistic Regression and

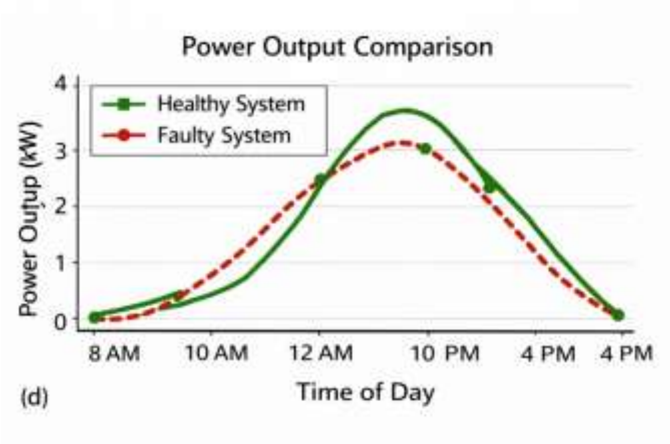
International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 03 | March - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

Decision Tree showed slightly lower accuracy. The evaluation metrics, including precision, recall, F1-score, and confusion matrix, confirmed the reliability of the models. However, when tested on data from different PV systems, performance decreased, indicating the need for system-specific model training.









6. Quantitative Security Analysis

This evaluates the security of a system using measurable metrics and numerical data. It involves identifying potential threats, vulnerabilities, and risks, and analyzing their impact using statistical and mathematical methods. Techniques such as probability analysis, risk assessment models, and performance metrics are used to measure system security. This approach helps organizations make informed decisions, prioritize security measures, and improve the overall reliability and protection of

7.CONCLUSION

This project demonstrates the effective use of machine learning, particularly the **Random Forest** algorithm, for detecting and classifying faults in photovoltaic (PV) systems using electrical and environmental data. Key parameters such as voltage, current, temperature, and irradiance are used to identify faults like partial shading and dirt accumulation that reduce solar power output. Data preprocessing and supervised learning techniques help improve model accuracy and reliability under the same system conditions used for training. Experimental results show that the Random Forest model achieves high accuracy and reliable fault classification. However, performance may decrease when applied to different PV systems, indicating the need for system-specific training for better real-world implementation.

8. FUTURE SCOPE

The scope of this project is to develop a machine learning-based framework for detecting faults in photovoltaic (PV) systems using electrical and environmental data. It focuses on identifying normal operation, partial shading, and dirt accumulation. Various supervised algorithms such as Logistic Regression, Decision Tree, Random Forest, SVM, and ANN are applied and evaluated using standard performance metrics. The study also examines limitations in model performance across different PV systems and emphasizes the need for system-specific training.

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.