e-ISSN:2582-7219

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 7, Issue 13, April 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.521

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| | Monthly Peer Reviewed & Refereed Journal |

|| Volume 7, Issue 13, April 2024 ||

International Conference on Intelligent Computing & Information Technology (ICIT-24)

Organized by

Erode Sengunthar Engineering College, Erode, Tamilnadu, India

Detection and Diagnosis of Faulty Photovoltaic Modules Through Machine Learning Algorithms

Ms. C. Kotteeswari[1], J.V.Swethaa[2], C.Premalatha[3], K. Tamilselvan[4]

Assistant Professor, Dept. of C.S.E, Velalar College of Engineering and Technology, Erode, Tamil Nadu, India[1]

Student, Dept. of C.S.E, Velalar College of Engineering and Technology, Erode, Tamil Nadu, India[2, 3, 4]

ABSTRACT: The rising number of solar power plants globally, driven by environmental concerns, faces maintenance hurdles, notably in detecting faulty photovoltaic (PV) modules, especially in larger or remote setups. Timely identification and replacement of these modules are crucial to mitigate potential serious incidents. This study employs machine learning to address the challenge of detecting and estimating electrical faults within PV farms, selecting logistic regression for its superior decision- making. It incorporates various parameters such as distributed PV current measurements, plant current, voltage, and power readings, along with temperature, radiation levels, and fault resistance. Thirty fault feature attributes, focusing on four major ones, aid fault detection through logistic regression. Recursive Feature Elimination (RFE) is crucial for feature selection before applying logistic regression, systematically removing irrelevant or redundant features to streamline the dataset and enhance efficiency. The study explores a range of fault types commonly found in PV farms encompassing string fault, string-to-ground fault, and string-to-string fault. Achieving a remarkable accuracy of 99.42% underscores the robustness and efficacy of the proposed machine learning methodology in detecting and estimating electrical faults within PV farms.

KEYWORDS: Photovoltaic (PV) fault detection, Recursive Feature Elimination, Logistic Regression, Solar Power Generation

I. INTRODUCTION

The worldwide transition towards eco-friendly energy solutions has driven the extensive uptake of photovoltaic (PV) technology for solar energy generation. Ensuring the reliable performance of PV modules is paramount for the efficiency and longevity of solar energy systems. Detecting and mitigating malfunctioning PV modules are crucial to maintaining optimal energy production and system reliability. Traditional monitoring methods often struggle to identify subtle or evolving issues within solar arrays. To address this challenge, integrating machine learning algorithms has emerged as a promising solution for early and accurate detection of malfunctioning PV modules. By leveraging advanced data analytics and pattern recognition, these algorithms can analyze extensive performance data, enabling proactive identification of anomalies and timely maintenance interventions. This convergence of solar technology and artificial intelligence represents a significant stride towards enhancing the efficiency and resilience of photovoltaic systems in the pursuit of sustainable energy solutions. PV power plants, pioneering contributors to renewable energy since 1982, have become a cornerstone of global renewable energy sources. By the end of 2019, installed PV plant capacity reached 627 GW worldwide. Faults within PV plants can disrupt operations, affecting service continuity and lifespan. These faults, caused by various factors, impact different components of PV systems, compromising efficiency and reliability. Electrical faults typically fall into three main categories: string faults, string-to-ground faults, and string-to- string faults. String faults are localized issues within specific strings of PV modules, caused by factors like module malfunction, partial shading, or wiring issues. String-to- ground faults occur when a string of PV modules connects unintentionally to the ground, posing significant safety hazards. String-to-string faults occur when unintended electrical connections form between two or more strings, often due to wiring issues or environmental factors. Addressing these fault types is critical for maintaining the performance and safety of PV systems.

II. RELATED WORKS

Grimaccia [1] and team introduced a novel approach for detecting defects in photovoltaic (PV) modules using unmanned aerial vehicles (UAVs). Their post-processing tool analyzes remote aerial images to aid operation and maintenance personnel in identifying faults, potentially enhancing plant performance and energy yield. The research, based on real plant data, explores the effectiveness of UAV- based image processing techniques in the renewable energy sector. Zhang [2] et al. developed a drone-mounted infrared thermography system to rapidly detect and locate fouling on large-scale PV panels, enhancing power efficiency. Their approach combines preprocessing and detection methods, showcasing practical value for PV maintenance. Xiaoxia [3] et al. implemented a deep learning solution utilizing aerial images for precise defect analysis in large-scale PV farms, improving efficiency and accuracy. Their approach surpasses conventional methods, facilitating effective asset inspection and maintenance. The algorithm's effectiveness in defect detection is evidenced through comprehensive evaluation, showcasing its potential for efficient maintenance of PV modules. Carletti [4] et al developed an automated flying system for fault detection in PV plants using UAV-mounted thermal cameras. Their model-based approach incorporates structural regularity and novel techniques for local and global hot spot detection, optimized for real-time onboard processing, ensuring high accuracy in anomaly detection for efficient thermographic PV inspection. Zhang [5] et al. designed a data- driven approach using Random Forests and XGBoost for wind turbine fault detection, enhancing availability and reliability. Their method, validated through simulations, demonstrates robustness across various turbine models and working conditions, outperforming traditional methods like support vector machines. Aghaei [6] et al. implemented infrared-based real-time monitoring for PV systems using digital image processing techniques. Their method, employing a mounted IR camera on a light UAS, accurately detects defects and degradation, proving its reliability for PV module inspection. Tsanakas [7] et al. utilized image processing and Canny edge detection on thermal measurements for fault diagnosis in PV modules, effectively detecting hot-spot formations indicative of specific defects, enhancing performance assessment. Ebner [8] et al investigated non-destructive methods like infrared thermography, electro-, and photoluminescence imaging for PV module quality control, emphasizing their efficacy in detecting defects. Zhang [9] et al. proposed an ICA-based method for surface defect detection in photovoltaic modules, enhancing anomaly detection by removing background structures. Deitsch [10] et al. automated defective PV cell detection in electroluminescence images using SVM and deep CNN, enabling continuous, highly accurate monitoring. Buerhop [11] et al. investigated the reliability of IR-imaging in assessing PV-plants under operating conditions, revealing various failure mechanisms and their impact on power output. Their analysis, involving 260 dismantled modules, confirms the effectiveness of infrared-mapping in detecting defects and assessing power loss in PV-plants. Pierdicca [12] et al. utilized a DCNN to automatically detect damaged PV cells using data from a drone with a thermal infrared sensor. Their approach, validated on the "Photovoltaic Images Dataset," shows effectiveness in addressing degradation, with precise evaluation metrics. Chen [13] et al. present an XGBoost classifier for DDoS attack detection in SDN-based clouds, achieving high accuracy and low false positive rates. Their method, validated using TcpDump data and Hyenae attack simulations, offers scalability and fast detection speed for SDN controllers. Li [14] et al. developed a customized CNN for efficient classification of lung image patches with interstitial lung disease, offering versatility for medical image tasks. Akram [15] et al. propose an improved outdoor thermography scheme for defect detection in PV modules, enhancing performance and reliability. Their method involves modulating PV module temperature and employing image processing for accurate defect localization, crucial for optimizing PV system efficiency. Yilmaz [16] and Ozcelik propose a modified Perturb-and-Observe algorithm to enhance the performance of Maximum Power Point Tracking (MPPT) in PV systems. Their approach aims to address oscillation issues, improve speed, and optimize power output under varying radiation conditions, with promising simulation results. Fazai [17] et al. employ bibliometric analysis to explore fault diagnosis in photovoltaic systems using artificial intelligence, revealing trends and impactful methodologies. Their innovative approach combines statistical methods with expert content analysis to identify key research directions and algorithms. Hussain [18] et al. propose a fault detection method for PV systems, combining neural networks and Sugeno fuzzy logic. By utilizing input variables like irradiance and temperature, their approach achieves 99.28% accuracy in identifying short-circuited modules and 99.43% accuracy in detecting disconnected strings. Zhao [19] et al. examine line- line faults in solar PV arrays, discussing challenges to overcurrent protection devices (OCPDs) caused by fault currents influenced by maximum-power-point tracking. They propose solutions to mitigate safety hazards and enhance system reliability. Friedman [20] et al. explore Bayesian network classifiers, particularly focusing on Tree Augmented Naive Bayes (TAN) as a method that outperforms naive Bayes while maintaining computational simplicity. Their evaluation compares these classifiers with C4.5 and wrapper methods for feature selection, using datasets from the University of California at Irvine repository. S. Menard [21] examines the practical application of logistic regression analysis, emphasizing the need to translate odds ratios into probabilities for clearer interpretation by both researchers and lay audiences. Through recent literature examples, the article demonstrates the process of deriving probability implications from odds ratios. Schapire [22] et al. propose an explanation for the effectiveness of voting methods, showing that boosting increases margins of training examples, leading to improved classification accuracy.

III. PROPOSED SYSTEM

The proposed system integrates machine learning techniques for detecting and estimating electrical faults in photovoltaic (PV) farms, addressing challenges in maintenance and risk mitigation. Leveraging Logistic Regression for its decision- making prowess, the system incorporates various parameters including distributed PV current measurements, plant voltage, and temperature. With a focus on fault feature attributes like string and ground faults, Recursive Feature Elimination streamlines data for efficient fault detection. A classification tree aids in identifying fault types, supporting prompt intervention. By continuously monitoring real-time data, the system ensures timely identification of anomalies, enabling swift maintenance actions to mitigate potential risks. Through optimization techniques and ongoing refinement, the system aims to enhance fault detection accuracy and contribute to the reliability and efficiency of solar power plants worldwide.

A. *Data Acquisition*

A 250-kilowatt photovoltaic (PV) power plant simulation generated a dataset for training and testing, focusing on fault scenarios. The dataset comprised 3500 instances with 30 features, including temperature, radiation, and fault resistance measurements ranging from 10°C to 35°C, 100 W/m2 to 1,000 W/m2, and 1 Ω to 2,000 Ω, respectively. Fault cases F1, F2, and F3 represented string fault, string-to-ground fault, and string-to-string fault, respectively, occurring at the 0.2-second mark in the 0.4-second simulation duration. Features included statistical summaries of current readings from each string's two ammeters, capturing current at both ends. Supplementary metrics like total average DC power, current, and voltage provided insights into plant performance and fault conditions, facilitating a comprehensive analysis of fault impacts on the PV system.

B. *Data Splitting*

Data splitting is pivotal in machine learning workflows, essential for assessing model performance and ensuring adaptability to unseen data. Initially, the dataset is divided into features (X) and the target variable (y). Using positional indexing, features are separated, leaving the last column as the target variable. Next, the dataset is split into training and testing sets, with 80% allocated for training and 20% for testing. This ensures both sets represent the original dataset adequately. The inclusion of a random state parameter guarantees reproducibility across different runs. Data splitting enables training the model on a portion of the data while reserving another portion for evaluation, facilitating robust model assessment and validation.

C. *Data Formatting*

Data formatting is crucial in data management, ensuring datasets adhere to specific standards. Tasks like reading and writing CSV files are common, especially in European sources where semicolons serve as delimiters. Maintaining data integrity during transformations is paramount, achieved through Python's CSV module for seamless conversion between formats. Pandas facilitates loading formatted data into DataFrames for downstream analysis. This process, integral to the data preprocessing pipeline, ensures data consistency and compatibility for subsequent analytical tasks, without delving into specific code implementation.

D. *Feature Selection*

Feature Selection is pivotal in machine learning, aiding in identifying relevant features for predictive modeling. Utilizing Recursive Feature Elimination (RFE) with a Random Forest Classifier, the code efficiently selects the most informative features. Initially, a Random Forest Classifier is instantiated as the base model, known for its robust performance with high- dimensional datasets. RFE, initialized with the desired number of features, ranks and selects features based on the classifier's predictive power. The code then retrieves feature rankings and identifies selected features, aiding in understanding their importance in predicting the target variable. Visualizing feature rankings offers insights into each feature's relative importance, aiding interpretation and model refinement. The final model is trained using selected features, enhancing computational efficiency and potentially improving generalization. Evaluating the

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| | Monthly Peer Reviewed & Refereed Journal |

|| Volume 7, Issue 13, April 2024 ||

International Conference on Intelligent Computing & Information Technology (ICIT-24)

Organized by

Erode Sengunthar Engineering College, Erode, Tamilnadu, India

model's accuracy on the test set validates the feature selection process, ensuring the model focuses on informative features for improved interpretability, efficiency, and performance. Fig 1. Shows the RFE Ranking

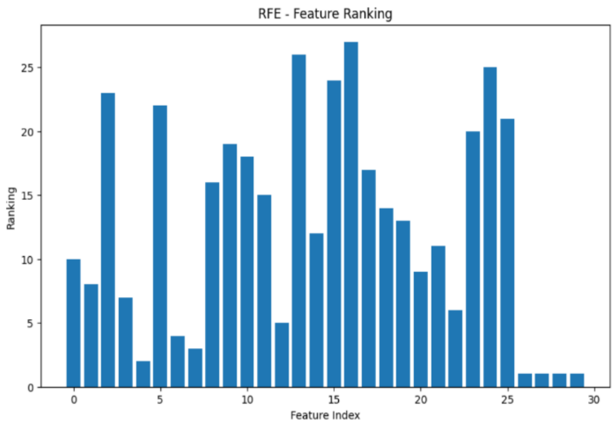

Figure 1 RFE Ranking

E. Logistic Regression-Based classification

Feature selection via Recursive Feature Elimination (RFE) with a Random Forest Classifier precedes training a Logistic Regression Classifier for binary classification. This approach yields a 99% accuracy on the test set, emphasizing the model's adeptness in predicting the target variable. By integrating feature selection techniques and logistic regression, the method constructs an interpretable and efficient classification model. Focusing solely on pertinent features enhances interpretability, efficiency, and potentially predictive accuracy, positioning logistic regression-based classification as a valuable tool across various machine learning tasks. Fig 2. Represents the classification tree.

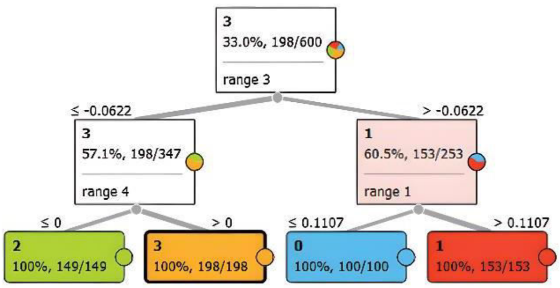

Figure 2: Classification Tree

In evaluating algorithmic performance, key metrics such as accuracy, precision, recall, and F1-score play a crucial role. Accuracy provides an overall measure of classification correctness, while precision and recall offer insights into an algorithm's ability to correctly identify positive instances and capture all positive instances, respectively. The F1-score synthesizes precision and recall into a single metric, providing a balanced assessment of the model's effectiveness. After a comprehensive evaluation, it is evident that Logistic Regression outperforms other algorithms, including SVM, AdaBoost, and KNN, across all evaluated metrics. Logistic Regression's superior performance in accuracy, precision, recall, and F1-score can be attributed to its ensemble learning approach and its ability to handle complex datasets efficiently. Fig 4 Performance Metrics

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| | Monthly Peer Reviewed & Refereed Journal |
|| Volume 7, Issue 13, April 2024 ||
International Conference on Intelligent Computing & Information Technology (ICIT-24)
Organized by
Erode Sengunthar Engineering College, Erode, Tamilnadu, India

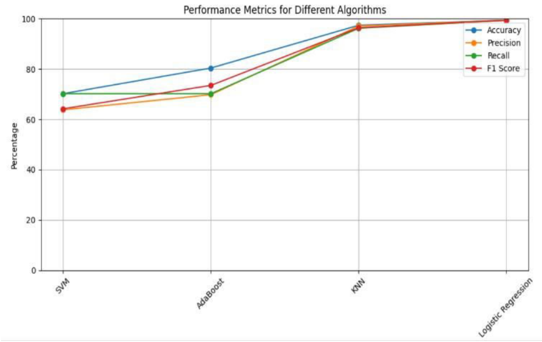

Figure 4: Performance Metrics

IV. Conclusion and Future work

In conclusion, the integration of machine learning algorithms, particularly logistic regression, offers a promising approach to enhance the detection of malfunctioning photovoltaic (PV) modules in solar power plants. By leveraging advanced data analytics and pattern recognition techniques, these algorithms enable the proactive identification of anomalies within PV farms, facilitating timely maintenance interventions to ensure optimal energy production and system reliability. The study emphasizes the importance of prompt identification and replacement of faulty modules to mitigate the risk of potential serious incidents and enhance the efficiency of solar power plants.

Moving forward, future research endeavors could focus on several aspects to further improve fault detection methodologies in PV systems. Firstly, comparative analyses of different fault detection algorithms could be conducted to determine the most effective approach under various operating conditions and fault scenarios. Additionally, research efforts could explore the development of novel fault detection techniques that address specific challenges, such as fluctuating irradiation and the presence of protection diodes in PV arrays

References

[1] F. Grimaccia, S. Leva, and A. Niccolai, “PV plant digital mapping for modules defects detection by unmanned aerial vehicles”, Aug.2017.
[2] P. Zhang, L. Zhang, T. Wu, H. Zhang, and X. Sun, “Detection and location of fouling on photovoltaic panels using a drone-mounted infrared thermography system”, Feb. 2017.
[3] X. Li, Q. Yang, Z. Lou, and W. Yan, “Deep learning- based module defect analysis for large-scale photovoltaic farms”, Mar. 2019.
[4] V. Carletti, A. Greco, A. Saggese, and M. Vento, “An intelligent ying system for automatic detection of faults in photovoltaic plants”, 2020.
[5] D. Zhang, L. Qian, B. Mao, C. Huang, B. Huang, and Y. Si, “A data driven design for fault detection of wind turbines using random forests and XGboost”, 2018.
[6] M. Aghaei, A. Gandelli, F. Grimaccia, S. Leva, and R. E. Zich, “IR real time analyses for PV system monitoring by digital image processing techniques”, 2015.
[7] J. A. Tsanakas, D. Chrysostomou, P. N. Botsaris, and A. Gasteratos, “Fault diagnosis of photovoltaic modules through image processing and canny edge detection on eld thermographic measurements”, 2015.
[8] R.Ebner, B.Kubicek and G.Ujvari, “Non-destructive techniques for quality control of PV modules: Infrared thermography, electro-and photoluminescence imaging”,2013.
[9] X. Zhang, H. Sun, Y. Zhou, J. Xi, and M. Li, “A novel method for surface defect detection of photovoltaic module based on independent component analysis”, 2013.
[10] S.Deitsch, V.Christlein, S.Berger, C.Buerhop-Lutz, A.Maier, F.Gallwitz, and C. Riess, “Automatic classification of defective photovoltaic module cells in electroluminescence images”, 2019.
[11] C. Buerhop, D. Schlegel, M. Niess, C. Vodermayer, R. Weimann, and C. J. Brabec, “Reliability of IR- imaging of PV-plants under operating conditions”, 2012.

[12] R. Pierdicca, E. S. Malinverni, F. Piccinini, M. Paolanti, A. Felicetti, and P. Zingaretti, "Deep convolutional neural network for automatic detection of damaged photovoltaic cells", 2018.
[13] Z. Chen, F. Jiang, Y. Cheng, X. Gu, W. Liu, and J. Peng, "XGBoost classifier for DDoS attack detection and analysis in SDN-based cloud", 2018.
[14] Q. Li, W. Cai, X. Wang, Y. Zhou, D. D. Feng, and M. Chen, "Medical image classification with convolutional neural network", 2014.
[15] M. W. Akram, G. Li, Y. Jin, X. Chen, C. Zhu, X. Zhao, M. Aleem, and A. Ahmad, "Improved outdoor thermography and processing of infrared images for defect detection in PV modules", 2019.
[16] M.Ozcelik and A.Yilmaz, "Improving the performance of MPPT in PV systems by modified perturb-and-observe algorithm" 2015.
[17] R. Fazai, K. Abodayeh, M. Mansouri, M. Trabelsi and G. E. Georghiou, "Machine learning-based statistical testing hypothesis for fault detection in photovoltaic systems", 2019
[18] M. Hussain, M. Dhimish, S. Titarenko and P. Mather, "Artificial neural network based photovoltaic fault detection algorithm integrating two bi-directional input parameters", 2020
[19] Y. Zhao, J. Palma, J. Mosesian, R. Lyons and B. Lehman, "Line-line fault analysis and protection challenges in solar photovoltaic arrays", 2013.
[20] N. Friedman, D. Geiger and M. Goldszmidt, "Bayesian network classifiers" ,1997.
[21] S. Menard, "Applied logistic regression analysis", 2001
[22] R. E.Schapire, Y. Freund, P. Bartlett and W. S. Lee, "Boosting the margin: A new explanation for the effectiveness of voting methods", 1998.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com