SVU-International Journal of Engineering Sciences and Applications (2024) 5(2): 140-152
Print ISSN 2785-9967 | Online ISSN 2735-4571
DOI 10.21608/SVUSRC.2024.279389.1198

Comparative Analysis of Machine Learning Techniques for Fault Detection in Solar Panel Systems

Montaser Abdelsattar[1,], Ahmed AbdelMoety[1], and Ahmed Emad-Eldeen[2]

Abstract: The utilization of Machine Learning (ML) classifiers offers a viable approach to improving diagnostic accuracy and system dependability in the pursuit of optimizing problem detection in solar panel systems. This work aims to conduct a thorough assessment of different Machine Learning (ML) classifiers in order to determine the most efficient models for detecting faults in solar panel systems. We rigorously tested and analyzed the classifiers AdaBoost, GaussianNB, Logistic Regression, Support Vector Classifier (SVC), Multi-Layer Perceptron (MLP), Decision Tree (DT), K-Nearest Neighbors (KNN), Random Forest (RF), and Extra Trees (ET). We evaluated the classifiers using their F1 scores, a crucial metric for measuring model performance in imbalanced class scenarios commonly encountered in fault detection tasks. The results show that the Decision Tree (DT), KNN, Random Forest (RF), and Extra Trees (ET) classifiers worked better than expected. All of them got perfect F1 scores of 1.000, which shows how well they can find bugs. On the other hand, AdaBoost demonstrated a lower F1 score of 0.591, suggesting possible constraints in its use for detecting faults in solar panel systems. This study advances fault detection in solar panels, enhancing system reliability and reducing maintenance costs. It also guides the development of sophisticated diagnostic tools, boosting solar technology adoption.

Keywords:. Artificial Intelligence; Fault Detection; Machine Learning; Predictive Maintenance; Solar Panel Systems.

1 Introduction

1.1 Background information

The current state of the global energy sector is experiencing a notable shift, as there is a growing demand for renewable energy sources. The need to address the environmental consequences associated with traditional fossil fuels and provide a sustainable energy trajectory for the future drives this shift [1]. Solar energy, which is captured by Photovoltaic (PV) systems, has become a fundamental component of renewable energy solutions owing to its ample accessibility and comparatively minimal ecological impact [2].

Solar panel systems use PV cells to transform sunlight into electrical energy [3]. Companies implement these systems at various scales, from small residential configurations to massive solar farms [4]. Solar energy presents a viable and environmentally friendly substitute for conventional energy sources [5]. However, the overall energy output and economic and environmental sustainability of solar panel systems heavily rely on their efficiency and efficacy [6].

Although solar panel systems offer numerous advantages, they are susceptible to a range of flaws and inefficiencies that can have a substantial impact on their overall performance [7]. The challenges encompass a spectrum of severity, spanning from simple concerns such as soiling and shading to more serious complications such as open-circuit faults, short-circuit faults, and inverter errors [8]. The quick detection and diagnosis of these defects are of utmost importance in order to uphold optimal operational efficiency and extend the lifespan of the system [9].

Historically, the identification of faults in solar panel systems has depended on physical examinations and regular maintenance assessments. Although these procedures are essential, they require a substantial amount of effort, consume a large amount of time, and frequently prove to be useless in detecting defects prior to their

Received: 25 March 2024/ Accepted: 27 April 2024
Corresponding Author Montaser Abdelsattar, Montaser.A.Elsattar@eng.svu.edu.eg
1. Electrical Engineering Department, Faculty of Engineering, South Valley University, Qena 83523, Egypt.
2. Renewable Energy Science and Engineering Department, Faculty of Postgraduate Studies for Advanced Sciences (PSAS), Beni-Suef University, Beni-Suef 62511, Egypt.

escalation into substantial efficiency losses or damage [10].

Machine learning (ML) has significantly transformed the domain of fault detection in solar panel systems in recent years. By utilizing the large quantities of data produced by these systems, ML algorithms can acquire the ability to recognize intricate patterns and irregularities that signify the existence of defects. The implementation of automated fault detection not only improves the precision and effectiveness of data but also leads to a substantial decrease in operational expenses [11].

1.2 Motivation behind the study

Historically, the identification of faults in solar panel systems has posed significant difficulties, frequently necessitating periodic human inspections that are both resource-intensive and incapable of promptly detecting concerns. The monitoring and diagnostics of solar energy systems require a more advanced approach due to their dynamic and complex character, as well as their unpredictability in external circumstances [9]. The inspiration for this study originates from the pressing requirement for effective, precise, and automatic fault detection techniques that can guarantee the optimal functioning of solar panel systems.

1. Improving Solar Panel Efficiency and Reliability: Through precise identification and diagnosis of problems, there exists the potential to greatly improve the performance and dependability of solar panel systems, thereby making a substantial contribution to the total efficiency of solar energy generation [9].

2. Reducing Maintenance Costs and Downtime: Using automated fault detection systems can effectively reduce potential problems, reducing the need for manual inspections, limiting system downtime, and lowering maintenance costs at the same time [12].

3. Advancing Renewable Energy Technologies: The objective of this study is to enhance the monitoring and maintenance procedures of different renewable energy technologies, thereby making a valuable contribution to the wider field of renewable energy management [13].

4. Fostering Sustainable Energy Solutions: The primary objective of this study is to contribute to the worldwide transition towards sustainable energy sources by enhancing the operational efficiency and dependability of solar power, which is widely regarded as one of the most promising renewable energy technologies [14].

1.3 The Problem Statement

The growing dependence on solar energy underscores the imperative of preserving the optimal functionality of solar panel systems. Conventional fault detection techniques frequently exhibit inefficiency since they lack the capacity for timely identification and necessitate substantial physical exertion. The present work aims to fill the existing research gap by examining the optimal utilization of ML techniques for the purpose of automated fault detection in solar panel systems. Although ML has the ability to significantly transform fault diagnostics by utilizing predictive analytics, there is a need for a comprehensive evaluation of several ML classifiers, such as Logistic Regression, Support Vector Classifier (SVC), Decision Trees (DT), Multilayer Perceptron (MLP), K-Nearest Neighbors (KNN), Random Forest (RF), Extra Trees (ET), AdaBoost, and Gaussian Naive Bayes. The objective of our study is to assess the efficacy of these classifiers in the identification and diagnosis of defects, with the intention of making a valuable contribution to the development of fault detection approaches and improving the precision and effectiveness of solar energy generation. Objectives of the Research:

1. Evaluate ML Classifiers: The goal is to compare and contrast how well different ML classifiers, like Logistic Regression, SVC, DT, MLP, KNN, RF, ET, AdaBoost, and Gaussian Naive Bayes, find problems in solar panel systems in a methodical way.

2. Determine Accuracy and Reliability: The objective is to determine the most accurate and reliable ML classifiers for defect identification and diagnosis in solar panel monitoring in order to assess their appropriateness for real-world applications [15].

3. Analyze Feature Importance: The goal is to find the features that give the best indications of system problems so that we can see how different features derived from solar panel systems affect the accuracy of each classifier [16].

4. Optimize Fault Detection Models: The objective is to investigate model optimization methods for the most effective ML classifiers, with the goal of improving their ability to detect faults while decreasing the occurrence of

False Positives (FP) and False Negatives (FN).

5. Enhance Renewable Energy Systems: With the aim of improving the effectiveness and long-term viability of renewable energy technologies, specifically solar panel systems, the utilization of ML techniques for sophisticated fault detection is proposed. The purpose of this objective is to decrease maintenance expenses and minimize system downtime while simultaneously promoting the wider acceptance and enhancement of renewable energy sources.

2 Methodology

2.1 Dataset

A dataset of operational data from solar panel systems supports this study. The dataset specifically focuses on performance indicators and environmental circumstances. The dataset consists of 3000 observations and includes data on current and voltage measurements from two independent strings (referred to as S1 and S2), light intensity measured in kiloLux, and ambient temperature in degrees Celsius. The aforementioned factors function as the principal characteristics for the ML models to examine and acquire knowledge from, with the ultimate objective of identifying faults within the solar panel systems.

Figure 1 displays histograms that illustrate the distribution of major operating parameters of the solar panel systems in the dataset, highlighting the variability and trends observed. The visualization displays the current output in amperes for strings S1 and S2 (a & b), voltage measurements in volts for S1 and S2 (c & d), as well as the light intensity in kiloLux (e) and ambient temperature in degrees Celsius (f). The presented histograms provide insights into the various operational conditions, encompassing the fluctuations in current and voltage that may indicate performance status or potential faults. Additionally, they demonstrate the influence of environmental factors such as light intensity and temperature, which are crucial for evaluating the energy production efficiency of solar panels across different scenarios.

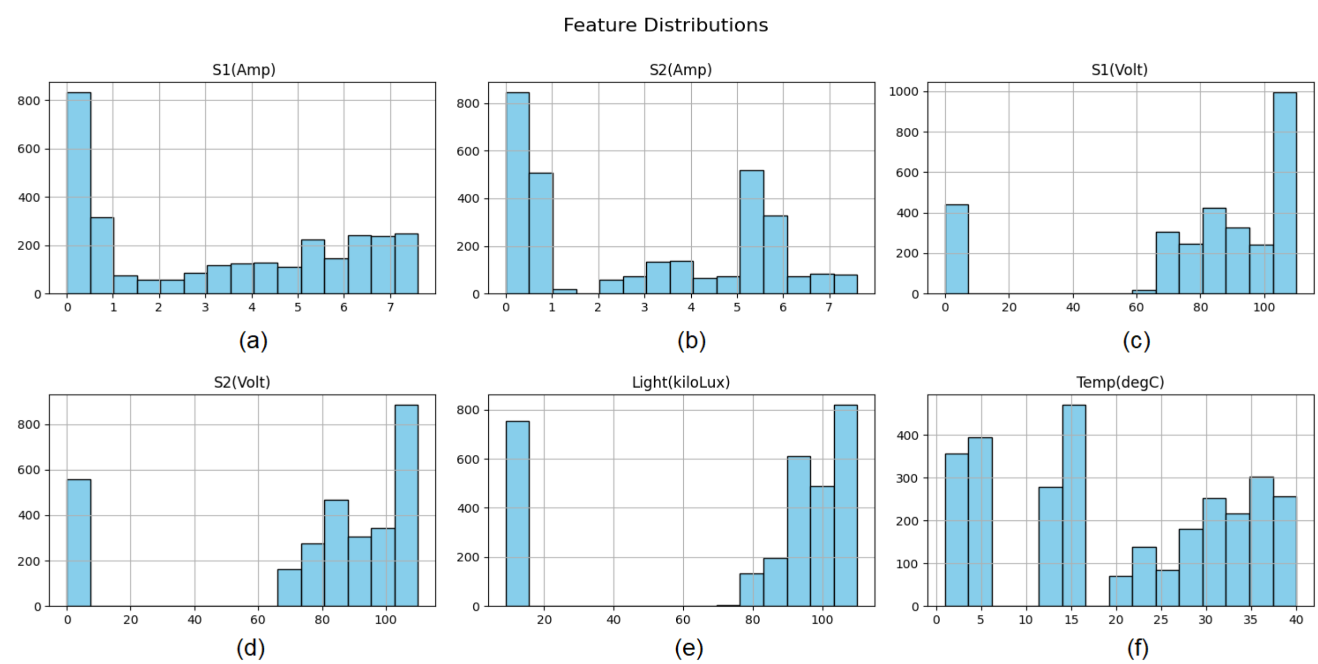

Fig. 1 Operational and environmental parameter distributions in solar panel systems: (a) S1(Amp), (b) S2(Amp), (c) S1(Volt), (d) S2(Volt), (e) Light(KiloLux), (f) Temperature(°C).

Figure 2 presents the statistical summary of the dataset, using boxplots to summarize the central tendency, dispersion, and outliers for each numerical feature. Using this picture makes it easier to find problems and better understand how the data is spread out, which are both important for the preprocessing and feature selection steps in fault detection models.

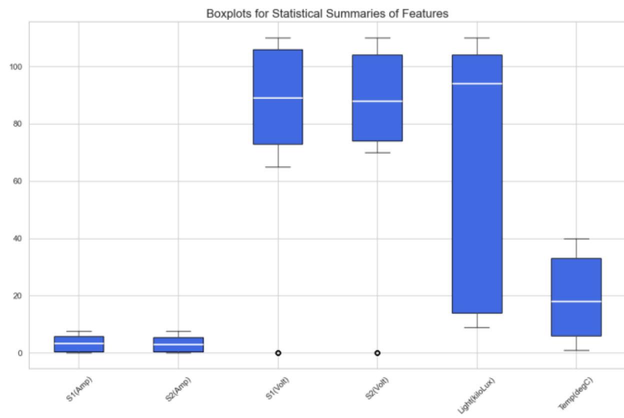

Fig. 2 Statistical summaries of solar panel parameters: central tendency, dispersion, and outliers.

Figure 3 visually displays the distribution of fault classes within a dataset from solar panel systems through a count plot. The visual representation, generated using the Seaborn library due to its visual attractiveness and interpretability, effectively displays a dataset that is uniformly distributed, consisting of 1,000 occurrences of 'Normal', 'Open', and 'Line-to-Line' fault states. The consistent allocation of data across different classes highlights the dataset's organized structure, enabling an equitable assessment of ML classifiers. Significantly, every bar is marked with the precise number of occurrences, augmenting the plot's informativeness.

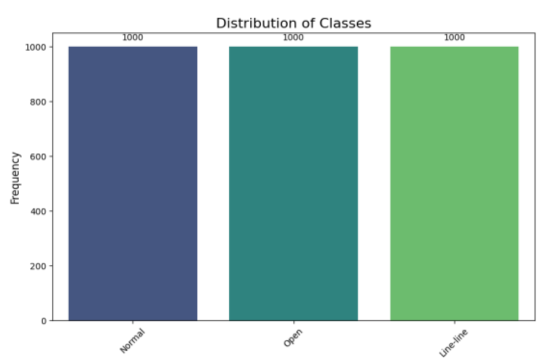

Fig. 3 Distribution of fault classes in solar panel systems.

Figure 4 illustrates the interrelationships among the features of the dataset, demonstrating the correlations between different parameters. The heatmap offers significant insights into the presence of multicollinearity and the relative strength of linear relationships among features. This information is crucial for directing the process of feature engineering and selecting suitable models for fault identification.

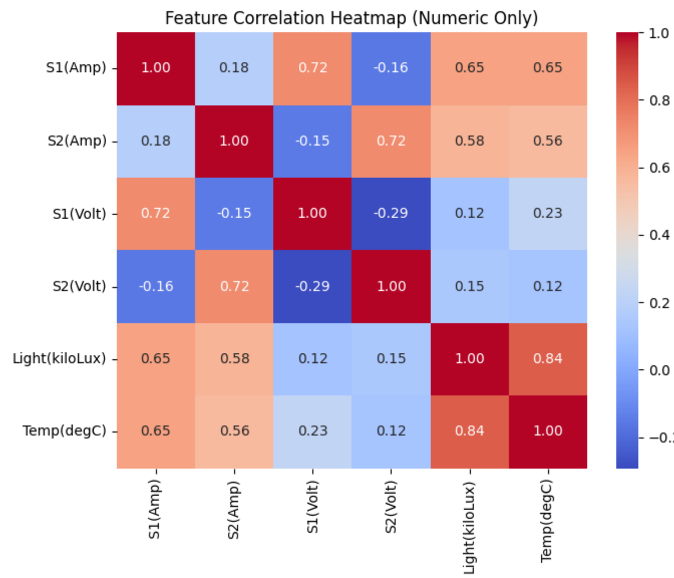

Fig. 4 Inter-Feature correlation analysis in solar panel systems: Insights into multicollinearity and associations.

We have carefully curated the dataset, incorporating encoded categorical variables like operational statuses for efficient analysis using ML techniques. We validate the models' performance using a split of 80% training data and 20% testing data, ensuring a thorough examination of their defect detection skills. The method of standardizing features by scaling highlights the preprocessing procedure aimed at normalizing the data in order to meet the requirements of various ML algorithms employed in this research.

2.2 Machine Learning

The present work utilizes a systematic methodology to assess the efficacy of different ML classifiers in identifying anomalies within solar panel systems. There are several essential processes involved in the methodology, including data collection and preprocessing, classifier setup and training, assessment, and analysis.

1. Data Collection and Preprocessing:

The dataset consists of 3,000 observations regarding the operations of a solar panel system. It includes seven features: measurements of current and voltage from two strings (S1 and S2), light intensity measured in kiloLux, and ambient temperature measured in degrees Celsius. The variable of interest signifies the fault status of the system, categorised into distinct groups.

The preprocessing stage involved encoding categorical variables using the LabelEncoder to transform them into a machine-processable format. We subsequently divided the dataset into two sets, a training set containing 80% of the data and a testing set containing 20%, to ensure a comprehensive evaluation of the classifier's performance. We applied feature scaling using StandardScaler to normalize the feature space, aiming to mitigate any potential bias in the classifiers' performance due to the magnitude of the features.

2. Classifier Initialization and Training:
This study chose a wide range of ML classifiers, such as Logistic Regression, SVC, DT, MLP, KNN, RF, ET, AdaBoost, and Gaussian Naive Bayes. To improve the performance of each classifier, particular parameters were initialized, such as'max_iter=2000' for logistic regression.

3. Evaluation:
We trained the classifiers on the preprocessed training data and then used them to generate predictions on the test set. We evaluated each classifier's performance using the F1 score, which is the harmonic mean of precision and recall. This metric strikes a compromise between the classifier's accuracy and its capacity to recall all pertinent instances. We produced confusion matrices to conduct a more comprehensive analysis of each classifier's performance, specifically focusing on evaluating the occurrence of FP and FN.

4. Analysis:
In addition to performance metrics, we computed and displayed the correlation matrix of classifier predictions to analyze the concurrence and disparities in prediction patterns among the classifiers. The objective of this analysis was to reveal insights into the behavior of the models and determine which classifiers have a tendency to produce similar predictions.

The study process is depicted in Figure 5, whereby the first stage involves loading the dataset. This dataset encompasses operational and environmental data that is relevant to solar panel systems. After collecting data, the categorical variables in the dataset, such as operational statuses, are encoded to make it easier to analyze computationally. This prepares the dataset for the next step of dividing it into training and testing sets. This division allows for a strong assessment framework, in which the model is trained on a specific portion of the data and then evaluated on a different portion to evaluate its ability to generalize.

The subsequent pivotal stage entails feature scaling, which entails normalizing the feature set to prevent any individual attribute from exerting excessive influence on the model as a result of its magnitude. This procedure is particularly vital for classifiers that are sensitive to the magnitude of features, such as logistic regression and SVC. The classifiers in question have been trained especially on the scaled data, whereas other classifiers have been trained on the original data in order to preserve their fundamental operational features.

Each classifier undergoes a separate training step, which includes Logistic Regression, SVC, DT, MLP, KNN, RF, ET, AdaBoost, and Gaussian Naive Bayes. We evaluate each model after the training process by calculating F1 scores and confusion matrices. These metrics provide valuable information regarding the precision, recall, and overall accuracy of the models in detecting faults within solar panel systems. We carefully store each classifier's predictions for future analysis. We compute a thorough correlation matrix after processing all classifiers, which clarifies the linkages and prediction patterns among the different models.

This particular stage holds significant importance in comprehending the relative performance and possible intersections in the defect detection algorithms employed by the classifiers. The last phases of the process focus on visualization, namely creating a correlation matrix to visually depict the relationships between classifiers. Additionally, we conduct a comparison of F1 scores across all models to underscore their efficacy in fault detection.

The methodology's flowchart highlights the comprehensive and comparative examination of ML techniques for improving the dependability and effectiveness of solar panel systems.

When doing a thorough examination of different ML algorithms for detecting faults in solar panel systems, it is essential to review existing works in the field. Table 1 presents a comprehensive analysis of the methodology, outcomes, and significant discoveries from pertinent research.

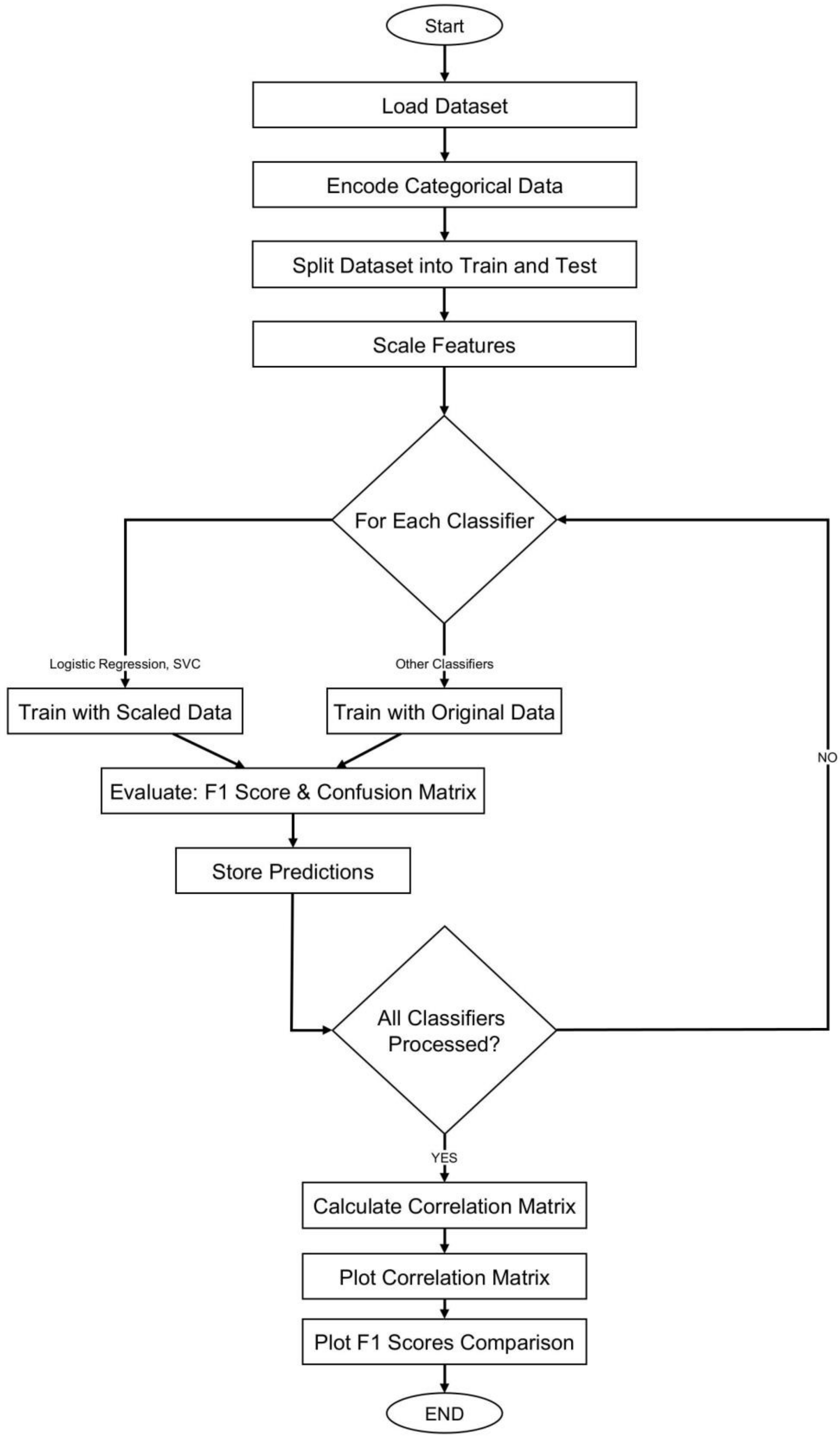

Fig. 5 Comparative analysis of ML classifiers for fault detection in solar panel systems: a methodological workflow.

Table 1 Comparative overview of ML techniques in solar panel fault detection studies.

Reference	Year	Focus of the Study	Key Findings	ML Techniques Used
[17]	2023	Internet of Things (IoT) and ML for monitoring, control, and fault detection in solar panels.	The system enhances operational efficiency and provides real-time fault detection.	ML, IoT.
[18]	2023	Comparison of various ML models to detect faults in solar panels based on IV curves.	RF found to be most effective for classification, robust against training dataset reduction.	RF, SMOTE (Synthetic Minority Oversampling Technique).
[19]	2023	Stacking-based ensemble learning for fault diagnosis in PV systems.	High accuracy in fault diagnosis (detection rate of 98.56%) using a stacking strategy of various ML algorithms.	Stacking-Based Ensemble Learning, Extra Tree Supervised Algorithm.
[20]	2023	Using physics-informed Deep Learning (DL) for fault detection in solar panel tracking systems.	Enhanced fault detection under diverse conditions, outperforming data-driven methods.	Physics Informed DL.
[21]	2022	Develops Ensemble Learning-based models for fault detection in grid-connected PV systems under uncertainty.	Improved diagnosis capabilities using interval kernel PCA-based Ensemble Learning classifiers.	Ensemble Learning, Interval Kernel PCA
[22]	2022	Applies ML models to assess performance and detect faults early in solar plants.	Achieved high accuracy and F-score using the J48 model.	ML (J48 model)
[23]	2021	Evaluation of various ML techniques for detecting anomalies and faults in PV systems.	Explored the application of traditional ML techniques like Isolation Forest for fault detection, providing insights into their suitability and performance trade-offs.	Isolation Forest, other traditional ML methods
[24]	2021	Development of a ML -based fault diagnosis system for power switching devices in grid-tied PV systems that operates online without additional sensors.	Demonstrated that the chosen ML technique, SVM, can reliably diagnose faults across a broad range of irradiance levels, efficiently and without increasing system complexity.	Support Vector Machine (SVM)
[25]	2021	Development of semi-supervised ML models for PV fault detection using a fraction of the labeled data typically required.	Demonstrated that positive unlabeled learning can effectively learn solar fault detection models and exceed the performance of fully supervised classifiers with minimal labeled data.	Positive Unlabeled Learning
[26]	2020	Combining Artificial Neural Networks (ANN) and fuzzy logic for fault detection in PV modules.	High accuracy in detecting short-circuited modules and disconnected strings in PV systems.	ANN, Sugeno Fuzzy Logic
[27]	2020	Introducing a new ML method using Reduced Kernel Partial Least Squares (RKPLS) for fault detection in nonlinear systems.	Reduced computation time and false alarm rates in fault detection.	RKPLS
[28]	2020	Developing an improved fault detection and diagnosis technique using Principal Component Analysis (PCA) and supervised ML classifiers.	Enhanced fault detection and diagnosis performance in PV systems under various operating conditions.	Supervised Machine Learning (SML) Classifiers
[29]	2020	Utilizing a Graph-Based Semi-Supervised Learning (GBSSL) method for fault detection and correction in PV arrays.	Effective identification and correction of faults, including unlearned conditions, in a range of environmental settings.	GBSSL
[30]	2019	Integrating ML with statistical testing for improved fault detection in PV systems.	Enhanced fault detection performance using Gaussian process regression and generalized likelihood ratio test.	Gaussian Process Regression, Generalized Likelihood Ratio Test
[31]	2019	Developing a ML -based fault classification system using thermographic images.	High training and testing efficiency, demonstrating improved performance over conventional techniques.	Artificial Neural Networks
[32]	2019	Semi-supervised learning algorithm using KNN for fault detection in electrical power systems.	Effective fault detection and classification with an accuracy of 97%.	KNN
[33]	2019	Using ML to determine the cleanliness of PV panels to optimize their efficiency.	Demonstrated robustness and efficiency under various classification algorithms.	Various ML classifiers
This Study		**Evaluating various ML classifiers to optimize fault detection in solar panel systems.**	**DT, KNN, RF, and ET achieved perfect F1 scores of 1.000, highlighting their effectiveness. AdaBoost scored lower (0.591), indicating potential limitations.**	**AdaBoost, GaussianNB, LR, SVC, MLP, DT, KNN, RF, ET.**

3 Results and Discussion

Figure 6 provides a detailed representation of confusion matrices for nine different classifiers in our thorough assessment of ML classifiers for defect detection in solar panel systems. The proposed models include Logistic Regression, SVC, DT, MLP, KNN, RF, ET, AdaBoost, and GaussianNB. These matrices offer a detailed understanding of the performance of each classifier, emphasizing the frequencies of true positive, false positive, true negative, and false negative. The comparison presented enables a thorough evaluation of the classifiers' efficacy in reliably detecting different fault states in solar panel systems, highlighting the inherent trade-off between sensitivity and specificity in each model. The visualization plays a crucial role in comprehending the various prediction behaviors and operational capabilities of these classifiers, hence providing valuable insights for choosing the most suitable model for practical implementation in the monitoring and maintenance of solar panel systems. From an analytical perspective, the research highlights the complex dynamics of problem identification using ML, providing a technique to improve diagnostic accuracy and system dependability.

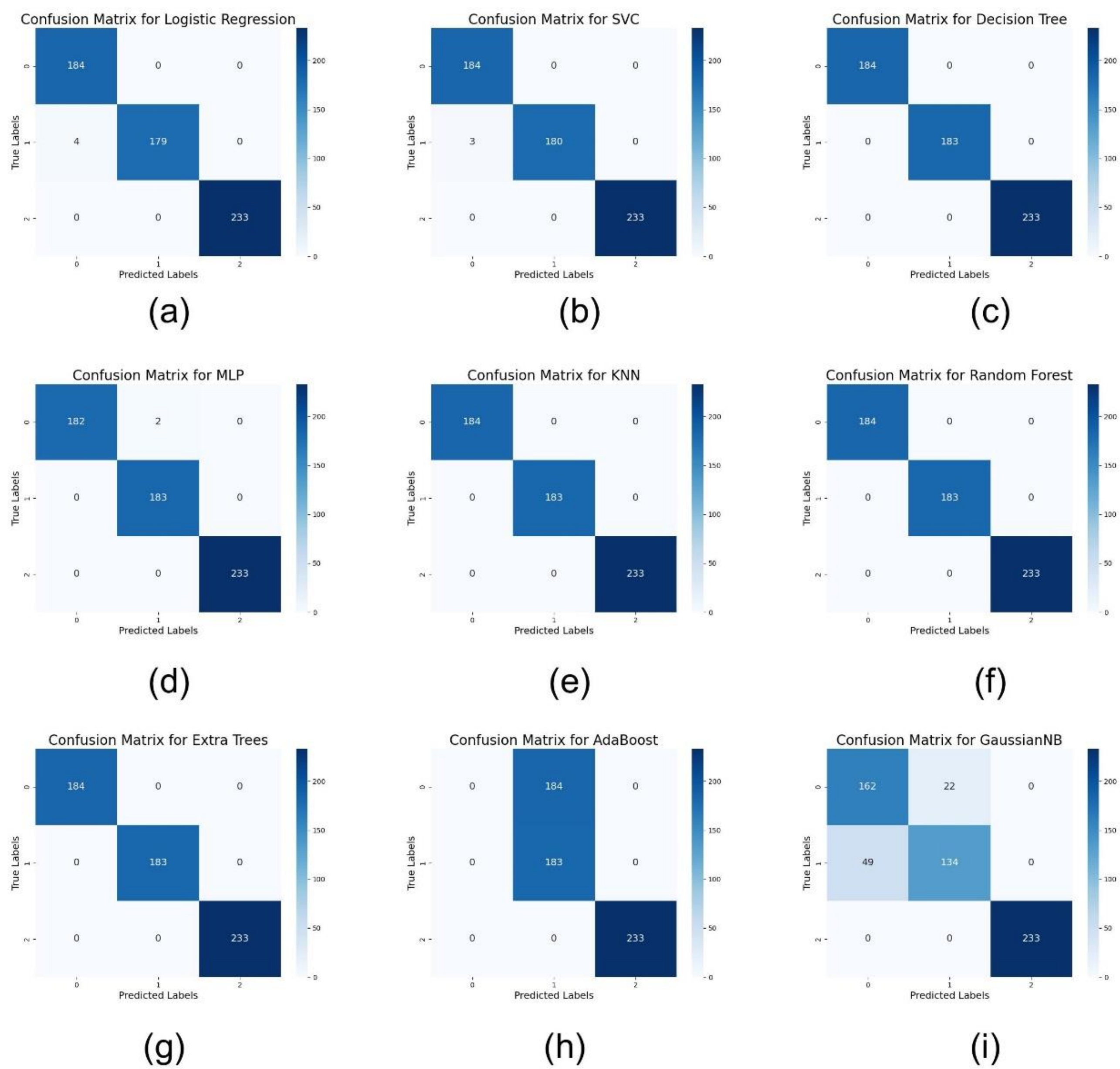

Fig. 6 Comparative analysis of confusion matrices for fault detection in solar panel systems across various ML classifiers: (a) Logistic Regression, (b) SVC, (c) DT, (d) MLP, (e) KNN, (f) RF, (g) ET, (h) AdaBoost, (i) GaussianNB.

In Figure 7, the complex interconnections and predictive congruence among several ML classifiers employed for the detection of problems in solar panel systems are revealed. This heatmap represents the level of concurrence among the classifiers, which include Logistic Regression and GaussianNB, in its predictions for defect identification. The diverse range of colors symbolizes the degree of association, with darker tones signifying greater agreement. The examination of these models is crucial in determining the similarities in their prediction patterns and the differences in their interpretations. This study provides insights into the various learning mechanisms and feature sensitivities present among different classifiers. Through the analysis of these correlations, the research not only brings attention to possible duplications in the behavior of classifiers but also identifies distinct contributions. This information may be utilized to inform the creation of ensemble methods or the choice of complementary classifiers for a reliable defect detection system. The research presented in Figure 7 provides a comprehensive analysis of model behavior within a collaborative framework, with the aim of enhancing the strategic integration of classifiers to achieve optimal performance in the diagnosis of solar panel faults.

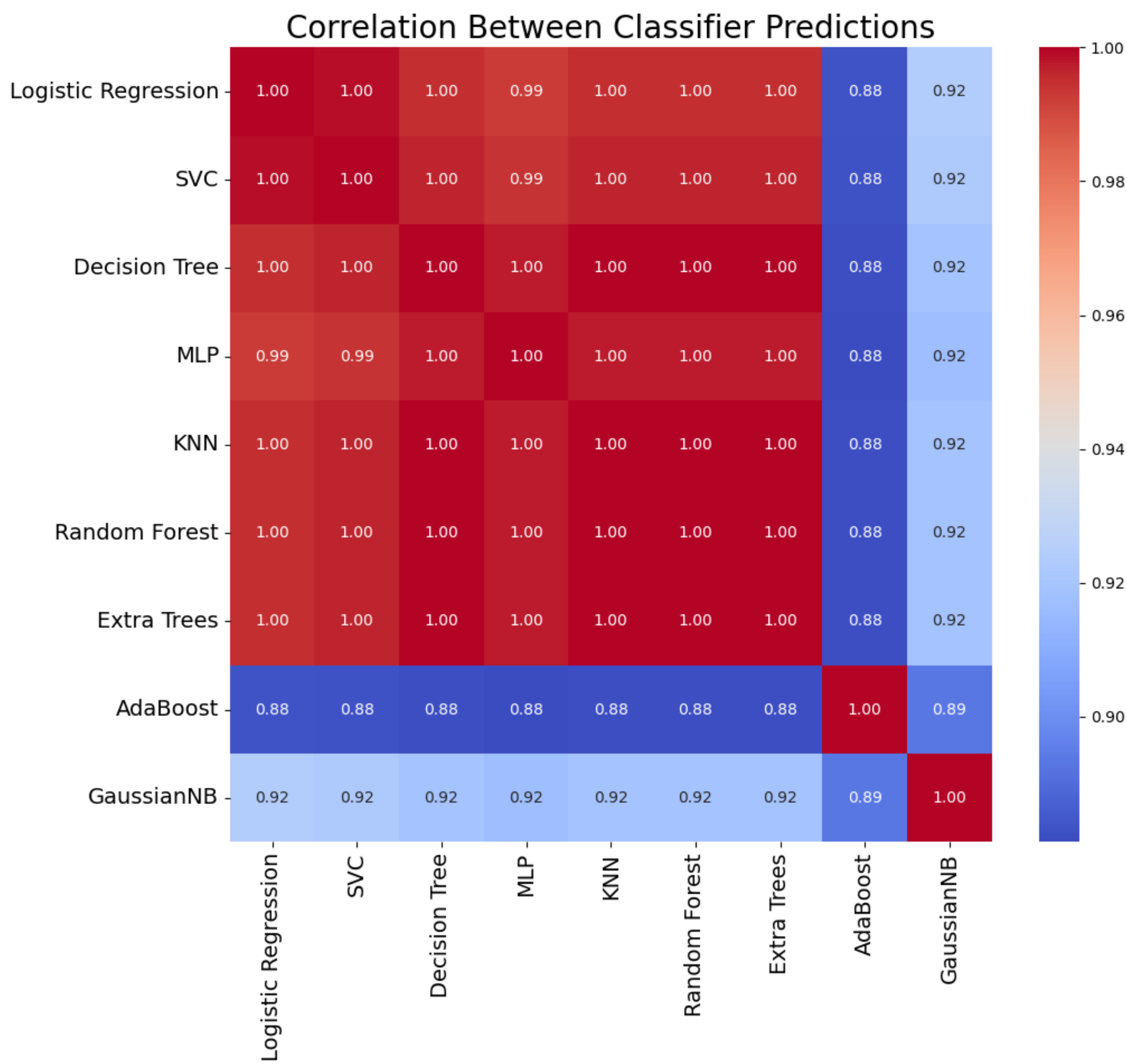

Fig. 7 Inter-Classifier prediction correlation analysis in solar panel fault detection.

In terms of fault detection within solar panel systems, Table 2 displays the efficiency ranking of different ML classifiers, as determined by their F1 scores. This research looks at how well nine different classifiers work. The classifiers include simple models like Logistic Regression and SVC as well as more complex ensemble techniques like RF and ET. We use the F1 ratings, adjusted to three decimal places for precision, as a standard to evaluate each classifier's efficacy. Ratings that are closer to 1.0 indicate a higher ability to detect faults. The application of boosting algorithms in this particular scenario presents hurdles, as evidenced by the lowest F1 score of 0.591 achieved by AdaBoost. In contrast, classifiers like DT, KNN, RF, and ET exhibit remarkable performance, achieving perfect scores of 1.000. Table 2 highlights the different levels of success achieved by various ML methods in detecting problems. This provides valuable information for choosing the most effective models to ensure the dependability and effectiveness of solar panel systems.

Table 2 Efficiency ranking of ML classifiers for fault detection in solar panel systems based on F1 scores.

Classifier	F1-Score
AdaBoost	0.591
GaussianNB	0.881
Logistic Regression	0.993
SVC	0.995
MLP	0.997
DT	1.000
KNN	1.000
RF	1.000
ET	1.000

In the evaluation of the effectiveness of ML classifiers in detecting faults in solar panel systems, the primary metrics employed are precision, recall, and the F1 score. The accuracy of a model in predicting positive observations is measured by precision Equation (1), which takes into account the ratio of properly predicted positive occurrences True Positives (TP) to the total number of positive forecasts produced, encompassing both right and wrong positive predictions TP and FP. The recall, as defined by Equation (2), evaluates the model's capacity to correctly identify all instances of positive instances. It quantifies the proportion of TP recognized out of all actual positive instances, which includes those that were missed FN. The F1 score, as represented by Equation (3), aligns these measures by computing their harmonic mean, resulting in a unified measure that encompasses both the precision and recall skills of the model. This metric holds significant importance in situations when there is an uneven distribution of classes, as it guarantees that both the model's ability to accurately predict positive cases and its capability to identify positive examples are taken into account. The application of the F1 score in assessing the efficacy of several classifiers in fault detection facilitates the determination of the optimal model for improving the operational efficiency and dependability of solar panel systems.

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{1}$$

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{2}$$

$$F_1 = \frac{2 \times \mathrm{Recall} \times \mathrm{Precision}}{\mathrm{Precision} + \mathrm{Recall}} \tag{3}$$

Our ML models for defect detection are evaluated using Precision, Recall, and the F1 score as the major performance metrics. These metrics are specified by Equations (1), (2), and (3) correspondingly. TP refer to fault conditions that have been accurately identified, whereas FP refer to normal conditions that have been wrongly identified as faults, and FN refer to faults that have not been detected by the model. Precision quantifies the level of accuracy of our model in correctly recognizing fault circumstances, whereas Recall evaluates the model's capability to recognize all pertinent occurrences of faults. The F1 score is a statistical measure that represents the harmonic mean of Precision and Recall. It is utilized to indicate the equilibrium between the two metrics and provides a single measurement for evaluating model performance in situations when the imbalance of classes may make Precision or Recall alone unreliable. These criteria are essential for assessing the dependability and usefulness of our classifiers for real-world implementation in solar panel systems, where the consequences of misidentification might be substantial. The mathematical model offered is characterized by carefully defined variables and their interrelations, which ensures clarity and precision.

The present study offers a thorough assessment of ML classifiers for identifying faults in solar panel systems. However, it recognizes specific constraints that open up opportunities for further research. An inherent constraint is the dependence on a static dataset, which might fail to encompass the complete spectrum of variations in real-life circumstances. Furthermore, the study did not include precise modeling of environmental factors that can impact the performance of solar panels, such as shadowing, bird droppings, and weather conditions. This omission may impair the applicability of the findings. Subsequent research endeavors could investigate the amalgamation of up-to-the-minute data and the simulation of environmental consequences in order to strengthen the reliability and resilience of the model. Furthermore, it is worth exploring the computational efficiency and scalability of the classifiers for large-scale solar farms. Investigating DL approaches and comparing their effectiveness with the classifiers described here would be a valuable pursuit to potentially reveal more intricate patterns in fault identification.

4 Conclusion

To summarize, this study conducted a comprehensive investigation of the effectiveness of different ML classifiers for detecting faults in solar panel systems. Our inquiry shed light on the range of performance among various classifiers, including Logistic Regression and Gaussian Naive Bayes, through a comparison analysis. The findings of the study revealed that specific classifiers, including DT, KNN, RF, and ET, have shown remarkable effectiveness in accurately detecting flaws. This highlights their potential for practical implementation in the monitoring of solar panel health.

Our findings clearly emphasized the need to carefully choose a ML model that is specifically designed to address the intricacies of defect detection tasks. These findings demonstrate the strength and flexibility of ML methods in guaranteeing the proper functioning of solar energy systems. Additionally, they highlight the significant influence of these technologies on the upkeep and long-term viability of renewable energy infrastructures.

Additionally, our examination of the predictive behaviors exhibited by the models highlighted the wide range of learning mechanisms they possess, establishing a strong foundation for future research in this field. These findings play a critical role in the progression towards the creation of hybrid or ensemble models that exploit the distinct capabilities of various classifiers in order to achieve enhanced diagnostic accuracy.

Optimizing the performance and dependability of solar panel systems through sophisticated defect detection procedures is a crucial task as global society progresses towards greener energy options. This study makes a useful contribution to the continuous endeavors aimed at tackling this difficulty by providing insightful viewpoints on the utilization of ML techniques to improve the effectiveness and dependability of solar energy.

We underscore the potential for innovation in integrating ML with renewable energy technology, hence promoting additional study in this domain. The utilization of interdisciplinary techniques plays a crucial role in enhancing operational efficiencies and promoting environmental sustainability and resilience in the context of global energy transformation.

Abbreviations

ANN	Artificial Neural Network
DL	Deep Learning
DT	Decision Tree
ET	Extra Trees
FN	False Negatives
FP	False Positives
GBSSL	Graph-Based Semi-Supervised Learning
IoT	Internet of Things
KNN	K-Nearest Neighbors
ML	Machine Learning
MLP	Multi-Layer Perceptron
PV	Photovoltaic
RF	Random Forest
RKPLS	Reduced Kernel Partial Least Squares
SMOTE	Synthetic Minority Oversampling Technique
SML	Supervised Machine Learning
SVC	Support Vector Classifier
SVM	Support Vector Machine
TP	True Positives

Declaration of competing interest

The authors affirm that there are no conflicts of interest related to the publication of this paper. They disclose no competing financial interests or personal relationships that might appear to have influenced the findings presented in this work.

Data availability

The data from this study can be obtained by requesting it from the corresponding author.

References

[1] D. Bogdanov et al., "Low-cost renewable electricity as the key driver of the global energy transition towards sustainability," Energy, 2021. [Online]. Available: https://doi.org/10.1016/J.ENERGY.2021.120467

[2] P. Choudhary and R. Srivastava, "Sustainability perspectives- a review for solar photovoltaic trends and growth opportunities," Journal of Cleaner Production, 2019. [Online]. Available:

https://doi.org/10.1016/J.JCLEPRO.2019.04.107.
[3] J. Briar and L. He, "A High Efficiency Solar Cell and System," in Proceedings of the 2021 IEEE 48th Photovoltaic Specialists Conference (PVSC), 2021. [Online]. Available: https://doi.org/10.1109/PVSC43889.2021.9518701
[4] A. Khanlari, A. Sözen, F. Afshari, and A. Tuncer, "Energy-exergy and sustainability analysis of a PV-driven quadruple-flow solar drying system," Renewable Energy, vol. 175, pp. 1151-1166, 2021. [Online]. Available: https://doi.org/10.1016/J.RENENE.2021.05.062
[5] A. Imthiyas et al., "Increasing the efficiency of solar panel by solar tracking system," IOP Conference Series: Materials Science and Engineering, vol. 993, 2020. [Online]. Available: https://doi.org/10.1088/1757-899X/993/1/012124
[6] Y. Jia, G. Alva, and G. Fang, "Development and applications of photovoltaic–thermal systems: A review," Renewable and Sustainable Energy Reviews, 2019. [Online]. Available: https://doi.org/10.1016/J.RSER.2018.12.030.
[7] A. Bazzi et al., "Fault impacts on solar power unit reliability," in Proceedings of the 2011 Twenty-Sixth Annual IEEE Applied Power Electronics Conference and Exposition (APEC), 2011. [Online]. Available: https://doi.org/10.1109/APEC.2011.5744749
[8] B. Nehme et al., "Faults Detection in PV Panels Using an Artificial Neural Network," in Proceedings of the 2023 IEEE 4th International Multidisciplinary Conference on Engineering Technology (IMCET), 2023. [Online]. Available: https://doi.org/10.1109/IMCET59736.2023.10368266
[9] A. Haque, K. Bharath, M. Khan, I. Khan, and Z. Jaffery, "Fault diagnosis of Photovoltaic Modules," Energy Science & Engineering, vol. 7, pp. 622-644, 2019. [Online]. Available: https://doi.org/10.1002/ese3.255.
[10] S. Madeti and S. Singh, "A comprehensive study on different types of faults and detection techniques for solar photovoltaic system," Solar Energy, vol. 158, pp. 161-185, 2017. [Online]. Available: https://doi.org/10.1016/J.SOLENER.2017.08.069.
[11] K. Jaskie, J. Martin, and A. Spanias, "PV Fault Detection Using Positive Unlabeled Learning," Applied Sciences, 2021. [Online]. Available: https://doi.org/10.3390/app11125599.
[12] S. Rao, G. Muniraju, C. Tepedelenlioğlu, D. Srinivasan, G. Tamizhmani, and A. Spanias, "Dropout and Pruned Neural Networks for Fault Classification in Photovoltaic Arrays," IEEE Access, vol. 9, pp. 120034-120042, 2021. [Online]. Available: https://doi.org/10.1109/ACCESS.2021.3108684.
[13] Y. Afridi, K. Ahmad, and L. Hassan, "Artificial intelligence based prognostic maintenance of renewable energy systems: A review of techniques, challenges, and future research directions," International Journal of Energy Research, vol. 46, pp. 21619-21642, 2021. [Online]. Available: https://doi.org/10.1002/er.7100.
[14] A. Bharatee, P. Ray, and A. Ghosh, "A Power Management Scheme for Grid-connected PV Integrated with Hybrid Energy Storage System," Journal of Modern Power Systems and Clean Energy, 2022. [Online]. Available: https://doi.org/10.35833/mpce.2021.000023.
[15] R. Fazai, K. Abodayeh, M. Mansouri, M. Trabelsi, H. Nounou, M. Nounou, and G. Georghiou, "Machine learning-based statistical testing hypothesis for fault detection in photovoltaic systems," Solar Energy, 2019. [Online]. Available: https://doi.org/10.1016/J.SOLENER.2019.08.032.
[16] E. Solano, P. Dehghanian, and C. Affonso, "Solar Radiation Forecasting Using Machine Learning and Ensemble Feature Selection," Energies, 2022. [Online]. Available: https://doi.org/10.3390/en15197049.
[17] S. Pranav et al., "An Integrated System for Monitoring & Control of Solar Panel using IoT & Machine Learning," in Proceedings of the 2nd International Conference on Modern Trends in Engineering Technology and Management, 2023. [Online]. Available: https://doi.org/10.21467/proceedings.160.53.
[18] M. Elgamal, A. Abdelmaksoud, and Y. Ismail, "Seamless Machine Learning Models to Detect Faulty Solar Panels," in 2023 3rd International Conference on Electronic Engineering (ICEEM), pp. 1-6, 2023. [Online]. Available: https://doi.org/10.1109/ICEEM58740.2023.10319482.
[19] A. Mellit, C. Zayane, S. Boubaker, and S. Kamel, "A Sustainable Fault Diagnosis Approach for Photovoltaic Systems Based on Stacking-Based Ensemble Learning Methods," Mathematics, 2023. [Online]. Available: https://doi.org/10.3390/math11040936.
[20] J. Zgraggen et al., "Physics Informed Deep Learning for Tracker Fault Detection in Photovoltaic Power Plants," in Annual Conference of the PHM Society, 2022. [Online]. Available: https://doi.org/10.36001/phmconf.2022.v14i1.3235.
[21] K. Dhibi et al., "Interval-Valued Reduced Ensemble Learning based Fault Detection and Diagnosis Techniques for Uncertain Grid-Connected PV Systems," IEEE Access, vol. PP, pp. 1-1, 2022. [Online]. Available: https://doi.org/10.1109/ACCESS.2022.3167147.
[22] E. Refaee, "Using Machine Learning for Performance Classification and Early Fault Detection in Solar Systems," Mathematical Problems in Engineering, 2022. [Online]. Available: https://doi.org/10.1155/2022/6447434.
[23] M. Ibrahim et al., "Machine Learning Schemes for Anomaly Detection in Solar Power Plants," Energies, 2021. [Online]. Available: https://doi.org/10.20944/preprints202112.0481.v1.
[24] Y. León-Ruiz et al., "Fault Diagnosis based on Machine Learning for the High Frequency Link of a Grid-Tied Photovoltaic Converter for a Wide Range of Irradiance Conditions," IEEE Access, vol. PP, pp. 1-1, 2021. [Online]. Available: https://doi.org/10.1109/ACCESS.2021.3126706.
[25] K. Jaskie, J. Martin, and A. Spanias, "PV Fault Detection Using Positive Unlabeled Learning," Applied Sciences, 2021. [Online]. Available: https://doi.org/10.3390/app11125599.
[26] R. Vieira et al., "PV Module Fault Detection Using Combined Artificial Neural Network and Sugeno Fuzzy Logic," Electronics, 2020. [Online]. Available: https://doi.org/10.3390/electronics9122150.
[27] M. Said, K. Abdellafou, and O. Taouali, "Machine learning technique for data-driven fault detection of nonlinear processes," Journal of Intelligent Manufacturing, vol. 31, pp.

865-884, 2020. [Online]. Available: https://doi.org/10.1007/S10845-019-01483-Y.

[28] M. Hajji et al., "Multivariate feature extraction based supervised machine learning for fault detection and diagnosis in photovoltaic systems," Eur. J. Control, vol. 59, pp. 313-321, 2020. [Online]. Available: https://doi.org/10.1016/j.ejcon.2020.03.004.

[29] H. Momeni et al., "Fault Diagnosis in Photovoltaic Arrays Using GBSSL Method and Proposing a Fault Correction System," IEEE Transactions on Industrial Informatics, vol. 16, pp. 5300-5308, 2020. [Online]. Available: https://doi.org/10.1109/TII.2019.2908992.

[30] R. Fazai et al., "Machine learning-based statistical testing hypothesis for fault detection in photovoltaic systems," Solar Energy, 2019. [Online]. Available: https://doi.org/10.1016/J.SOLENER.2019.08.032.

[31] V. Kurukuru et al., "Fault classification for Photovoltaic Modules Using Thermography and Machine Learning Techniques," in 2019 International Conference on Computer and Information Sciences (ICCIS), pp. 1-6, 2019. [Online]. Available: https://doi.org/10.1109/ICCISCI.2019.8716442.

[32] P. Wasnik, N. Phadkule, and K. Thakur, "Fault detection and classification based on semi-supervised machine learning using KNN," in 2019 International Conference on Innovative Trends and Advances in Engineering and Technology (ICITAET), pp. 79-83, 2019. [Online]. Available: https://doi.org/10.1109/ICITAET47105.2019.9170220.

[33] W. Hanafy, A. Pina, and S. Salem, "Machine Learning Approach for Photovoltaic Panels Cleanliness Detection," in 2019 15th International Computer Engineering Conference (ICENCO), pp. 72-77, 2019. [Online]. Available: https://doi.org/10.1109/ICENCO48310.2019.9027402.