Journal of Operation and Automation in Power Engineering

Vol. 14, No. 4, Dec. 2026, Pages: 275-285

http://joape.uma.ac.ir

Research Paper

Partial Shading Detection of Solar Panels Using Ensemble Bagged Trees Algorithm

Poorva Sharma and *Anamika Yadav* *

Department of Electrical Engineering, National Institute of Technology, Raipur, CG, India.

Abstract— *The adoption of photovoltaic (PV) energy is growing rapidly, leading to the installation of numerous solar power plants to meet rising electricity demand. Among the various operational challenges, partial shading significantly reduces the power output of PV panels. During routine maintenance, panels are typically cleaned to remove debris such as dust, dirt, or bird droppings—common causes of shading. However, shading severity varies across panels, especially in large PV installations, making uniform cleaning inefficient. To address this, the paper proposes a machine learning-based methodology for detecting and quantifying partial shading at the panel level. By analyzing power output, irradiance, and ambient temperature data, the proposed approach identifies the panels most affected by shading and classifies the severity into low, medium, and high categories. This enables smart maintenance prioritization and early fault prediction, preventing issues such as hotspots and module degradation while reducing operational costs. In a comparative study of eleven Machine Learning models, the ensemble bagged trees algorithm achieved the best performance 90% accuracy in identifying the most affected panels and 93.5% accuracy in classifying shading levels. The proposed solution is well-suited for real-time deployment in solar parks, offering an effective tool for predictive maintenance and optimized plant operations.*

Keywords—Partial shading, photovoltaic system, machine learning technique, modeling, simulation.

1. INTRODUCTION

Recognizing the world's future energy needs, governments and industries are increasingly investing in renewable energy to meet demand and explore alternative energy sources. It has been reported in [1] that renewable energy sources actively reduce CO^2 emissions and significantly decrease emission-related costs. By the end of 2023, India had generated 70,000MW of solar power, making it the dominant contributor to the country's renewable energy mix [2] and the cumulative global installed capacity has exceeded 760 GW [3, 4]. Solar panels actively capture sunlight to extract photovoltaic energy, playing a crucial role in solar power generation. Nonetheless, the variability of solar output and its integration with the main power grid pose several operational challenges [5]. Additionally, since solar panels are installed outdoors, they are susceptible to the accumulation of debris on their surface. This dirt and dust build-up, which is one of the reasons of partial shading, can severely reduce the panel's efficiency and its performance over time [6].

1.1. Motivation

It has been reported that partial shading can lead to a 20-40% decrease in the power output of solar panels, as it disrupts the amount of sunlight reaching the cells [7]. The effect of shading in the PV curve is mentioned in Fig. 1, where it is visible and it becomes evident that the power output of PV panels under shaded conditions gets reduced. Also, Shading triggers the formation of hotspots and contributes to panel degradation, which may ultimately lead to permanent damage [8]. To ensure optimal performance of PV panels, it is essential to reduce shading caused by dirt accumulation. The standard approach to address this issue involves cleaning the panels at regular intervals [9]. Also, there can be unequal deposition of debris leading to unequal shading in different panels for various reasons (e.g. bird droppings, solid dust accumulation, snow covering, leaves, passing clouds, shadows of nearby building structures, etc.) thereby posing a challenge in the maintenance activities, especially for large solar parks. Therefore, to carry out maintenance effectively, this paper proposes prioritizing the cleaning of solar panels with the highest shading levels. For this purpose, a smart system can be designed for the maintenance of solar panels from soiling and partial shading using priority-based cleaning of the panels which can detect the shading extent of each panel and identify the panels with highest shading levels. Thereafter, the panel having the highest shading should be given highest priority for the maintenance. As mentioned in [9], analyzing the shading level is also crucial from an economic perspective. This paper's analysis offers crucial insights that can help assess whether immediate maintenance is necessary or if there is ample time to implement a more flexible schedule. Achieving this requires accurately assessing the level of permanent shading, specifically by classifying panels as heavily, moderately, or lightly shaded based on the amount of debris accumulated on their surfaces. The proposed methodology can be particularly beneficial in cases where a large number of panels are mounted for mass generation of electricity.

Received: 25 Jun. 2024
Revised: 23 Aug. 2025
Accepted: 29 Aug. 2025
*Corresponding author:
E-mail: ayadav.ele@nitrr.ac.in (A. Yadav)
DOI: 10.22098/joape.2025.15344.2172

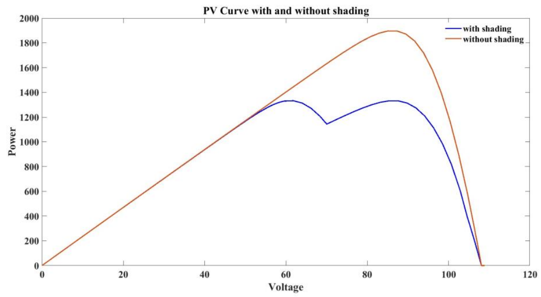

Fig. 1. PV curve with and without shading.

1.2. Literature review

There have been many studies in recent years regarding PV system monitoring and detection of various faults and maintenance of PV systems. The key challenges, present approaches, and opportunities for the predictive maintenance of PV systems are illustrated in [10]. The paper reviews predictive maintenance strategies for PV systems, noting that advanced methods like real-time sensing and machine learning offer higher fault detection accuracy but require greater investment compared to simpler inspection techniques. It also outlines the potential of integrating diverse approaches to overcome cost, scalability, and accuracy challenges, paving the way for more reliable and efficient PV system operation. Partial shading in PV systems provides valuable insights and practical solutions to enhance energy yield and system reliability through array reconfiguration and predictive modeling [11]. Different shading scenarios have been investigated in [12] in order to analyze its impact on the efficiency of the system. The study demonstrates that shading significantly reduces PV energy output with the impact strongly dependent on obstacle position. The detrimental impact of partial shading on PV systems has been underscored in [13] and suggestive methods thereof to prevent or mitigate the effect of partial shading have been further mentioned in the research article. However, this study mainly considers partial shading but does not deeply account for other influencing factors like temperature variation, dust deposition, or seasonal changes, which also affect PV performance. A smart PV monitoring system that leverages artificial neural networks is proposed in [14] to estimate output power from variables like irradiance and temperature, enabling the detection of shading and related faults. Paired with an IoT platform for real-time data visualization and alerts, this solution helps improve energy yield and reduce maintenance efforts. However in [14], only partial shading detection has been discussed without fault localization. For different shading strengths, the percentage by which the output of the solar panels gets affected was reported by [15]. It has been stated that under identical shading conditions, power loss is more severe in series-connected PV setups, while parallel connections experience relatively minimal reduction, emphasizing how electrical configuration can influence shading resilience. The monitoring technologies focused on various data processing modules and data transmission protocols are presented in [16]. The study also outlines persistent issues like handling massive data streams, overcoming communication interference, sustaining long-distance connectivity, and safeguarding system security, while suggesting that future advancements should focus on building monitoring systems that are more reliable, scalable, and secure. [17] Proposes a machine learning–based correction approach that detects shading effects caused by stationary obstacles such as buildings or trees and improves day-ahead PV power forecasts by leveraging inverter readings and prior-day irradiance data. While effective, the paper doesn't delve into how the model performs under real-world, continuously varying shading in diverse environments, indicating a gap in evaluating adaptability to dynamic or naturally occurring shading scenarios beyond the controlled setups used in the study. A method using digital image processing for binary partial shading detection on PV panels is presented in [18] with a detection accuracy of about 94.7% yet it stops short of evaluating how the method performs under diverse real-world conditions, like varying environments or integration with live power data. [19] presents reconfiguration strategies for PV arrays to overcome for the power losses caused due to partial shading highlighting that dynamic, metaheuristic-driven methods generally outperform static approaches, yet practical deployment remains constrained by complex wiring, high hardware demands, and scalability limitations. To compensate for the power mismatch losses, different PV array configurations have been proposed for 4 X 4 array size in [20] , however, the research is limited by its focus on a single array size and simulated scenarios, leaving open the question of how well these configurations perform across larger arrays or in varied real-world conditions. In [21], different photovoltaic array configurations are evaluated under partial shading, finding that certain arrangements (notably Total Cross Tied) maintain higher power output and lower mismatch losses compared to simpler layouts. However, the analysis is limited to simulated conditions and doesn't address performance in real-world, dynamically changing shading environments, leaving the real-life effectiveness of these configurations unverified. In [22], binary firefly algorithm-based array reconfiguration algorithm is discussed for fault detection achieving around 99% accuracy but the mentioned approach cannot be implemented on the already existing installed solar systems as it would require reconfiguration of the entire array system. Similarly, in [23] presented as solution for optimization of PV array under partial shading condition using Dynamic Leader-Based Collective Intelligence (DLCI) algorithm yet it remains untested in practical, hardware-based environments, leaving its real-world feasibility unproven. [24] suggests a Multi-Step Optimization Algorithm (MSOA), which is a dynamic reconfiguration strategy intended to rebalance the shading effects on the photovoltaic array yet it leaves unaddressed its performance in larger-scale or dynamic real-world settings. As a mitigation strategy, [25] has proposed a voltage equalizer to defend the PV system against partial shading achieving over 99% conversion efficiency with significantly fewer switches and sensors however, it leaves untested how the system performs in long-term real-world deployments. [26] introduced a method to quantify PV shading effect based on I-V curves when partial shading occurs using pixel-based comparisons between shaded and normal conditions. However, it does not demonstrate how robust or transferable this approach is when exposed to varying environmental conditions. [27] proposed a model-based method for detection of partial shading which is cost effective as it doesn't require expensive infrastructure, however, it is important to identify the precise location of faulty panels, after the detection has been accurately done. [28] proposes a sensorless, IoT-based method that uses semi-supervised learning to diagnose faults at the panel level in solar PV arrays while minimizing hardware requirements while leveraging both labeled and unlabeled data for effective fault detection. However, it does not investigate how the method performs across diverse environmental conditions or real-world installations. Diagnosis of partial shading and quantification of severity of shading has been discussed in [29], based on derivative characteristics and power deviations. The proposed I–V reconstruction method focuses solely on diagnosing and quantifying partial shading but does not integrate with any real-time control or reconfiguration system. In [30], a study on enhancing the efficiency of Linear Fresnel Reflector Solar Concentrators (LFRSCs) has been carried out using advanced optical modeling and optimization techniques. This work demonstrates that with robust simulations along with integration of advanced optical designs, LFRSC efficiency can be improved significantly. In the review paper [31], it has been summarized that use of super-hydrophobic transparent coatings in solar energy systems as a solution to mitigate dust accumulation and increase panel's efficiency looks beneficial but there is a need to develop standardized protocols to assess these coatings' longevity and performance in practical settings. Similarly, authors of [32] have also discussed and reviewed the application of using super-hydrophobic coatings for PV systems to mitigate the effects of dust accumulation. However, it has been stated in the article that the challenge of durability, long-term performance and cost-effectiveness remains. [33] Proposed the Diagonally Dispersed Total Cross Tied (D-TCT) Configuration, which rearranges PV modules diagonally within the array. To maximize the performance of PV systems under varying conditions like partial shading, the authors of [34] created a sophisticated Fuzzy Logic Controller based MPPT algorithm. Authors of [35] investigated about optimal planning for solar modules by strategically placing them to maximize energy efficiency to reduce shading effects. A novel approach to improve the panel's efficiency under partial shading has been presented in [36] by replacing conventional bypass diodes with electromagnetic relays. However, this approach requires modifications in the existing panel architecture and also, relays are bulkier than diodes which may further cause integration issues with PV panels. [37] carried out experiments to present a method for enhancement of PV array performance under conditions of partial shading by employing a Chaos Map-based physical reconfiguration strategy. The investigation regarding the effect of orientation of PV panels (portrait vs landscape) under partial shading scenarios has been demonstrated in [38]. Research article [39] underscores the importance of maximizing PV panel tilt and shading conditions to improve energy efficiency and reduce costs in building applications across a range of European environments. [40] investigates different methods like introduction of Magic Square View technique, a physical rearrangement strategy within the Total Cross Tied (TCT) configuration, with the aim to distribute shading effects more evenly across the array but the study does not address installation complexity, wiring modifications, or cost-benefit analysis for implementing the MSV configuration in practice. [41] introduced Modified Bridge-Linked (Modified BL) Configuration whose interconnection of solar modules lessens the negative effects of shading on the array as a whole but the analysis does not address how well the modified Bridge-Linked (BL) configuration scales across different PV array sizes—from residential installations to utility-scale farms.

Most of the previous researches have worked on the detection of partial shading or enumerating methods which can reduce the effect of partial shading but characterization of partially shaded panels on the basis of shading severity by the identification of the most affected panel as well as identification of the level of severity for timely mitigation has not entirely been covered in prior studies.

1.3. Paper contributions

In this paper, a very simple yet effective approach for identifying the shading phenomenon in PV systems has been discussed, which involves recognizing the panel impacted the most by partial shading and determining the shading level by using the power output value of each panel, irradiance, and ambient temperature values through machine learning. The proposed method is designed for implementation in solar parks and can be highly beneficial for the predictive maintenance of PV array systems, enabling timely interventions and reducing maintenance costs. Since the output of solar panels varies with changes in irradiance and temperature, it is often challenging to distinguish shading conditions from normal operating conditions, especially under low irradiance. Thus, machine learning techniques can be particularly advantageous for such cases as they have proficiency in identifying different patterns within large datasets [42, 43]. The major contributions of this work can be summarized as below:

1) The proposed methodology enables precise identification and localization of photovoltaic panels that are significantly impacted by shading effects. This capability facilitates proactive maintenance and fault diagnostics, particularly for issues such as hotspot formation and module degradation, which can critically impair overall system performance by reducing energy yield and operational lifespan.
2) Quantification of shading level with three definite demarcations as low, medium and high level of shading has been demonstrated in the paper. This work can significantly assist in predicting the impact of shading on efficiency, estimated power loss and scheduling of maintenance for solar panels based on the severity of shading level.
3) With the use of ensemble bagged trees machine learning algorithm, the detection of the most affected panel through shading and the identification of the level of shading have been done with an accuracy of 90% and 93.5% respectively.
4) Eleven machine learning algorithms have been included in a comparative study to show how the ensemble bagged trees algorithm outperforms the other approaches for the aforementioned techniques.

Thus, by using real-time data, this work can assist in determining the schedule for maintenance activities of the panels by identifying the panel with maximum shading and also the level of shading, henceforth using priority indicators for further actions to be taken. Fig. 2 presents the flowchart, which can be practically implemented for predictive maintenance by following the sequence shown. The process begins with identifying the panel subjected to the highest shading, followed by determining the shading level. Based on these two indicators, a decision can then be made regarding the need for immediate maintenance action. This can be done with the use of priority indicators for an automated system. Additionally, this information can help in predicting faults where shading is the primary cause.

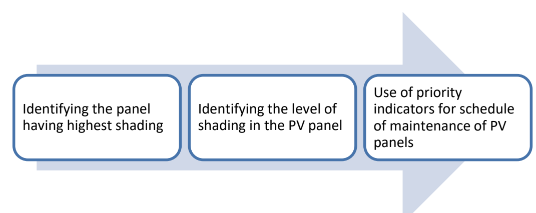

Fig. 2. Flowchart of predictive maintenance methodology.

2. METHODOLOGY: EXPERIMENTATION AND DATASET GENERATION

2.1. PV system details

This paper identifies two objectives that guide the adoption of an appropriate work sequence for priority-based maintenance of PV systems affected by shading. For this purpose, a simulation model is developed with two PV systems in which the first system represents a physical PV system under shading conditions, while the second system is its replica operating under ideal, non-shaded conditions. Each module's voltage and current values are measured, and thus, the power generated by each panel is measured for different irradiance and temperature values. The Power output of the shaded system is then compared with a non-shaded/ideal system under the same conditions i.e. solar irradiance and temperature. The power outputs and relative errors between the shaded and ideal systems are used as inputs/indicators for the classification learner algorithm. This workflow has been depicted in Fig. 3. By using only the power values and the relative error of the power difference as indicators, the following classification is performed:

1) Identification of the highest shaded panel
2) Level of shading with ranges defined as 0-33%, 34-66%, 67-100%

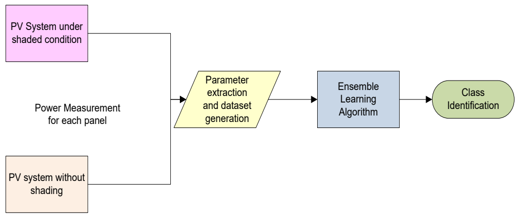

Fig. 3. Overview of proposed methodology workflow.

2.2. Ensemble bagged trees algorithm

Among the different classification techniques of machine learning, the ensemble bagged tree algorithm has an advantage when we are dealing with a complex dataset. Bootstrap sampling and decision trees are combined in ensemble bagged trees algorithm to lower model variance and increase accuracy. This learning merges the insights received from various learning models or base learners and gives the output. Through this process, the output received can have greater accuracy, and thus, improved decisions can be made. This machine-learning method uses several different decision trees instead of a single decision tree [14]. Thus, a group of weak learners combine and become a strong learner. Under the Bagging/ Bootstrap Aggregation, various subsets are taken randomly from the training dataset. Thereafter, each subset is given to train their decision tree [44]. The predictions made from each model/decision tree are used for the final prediction by the voting method. The purpose of choosing high-variance base learners (usually deep, unpruned trees) in practice is to maximize the benefit of averaging them. Thus training a bagged trees involves three primary steps i.e., i) bootstrap sampling, ii) tree induction which includes that on each bootstrap sample, a decision tree has to be trained by growing each tree to its full depth without pruning because deep trees have a high variance that the ensemble can average out and iii) aggregation, which is done for making a prediction on a new data by aggregating the outputs by majority vote to make the final prediction.

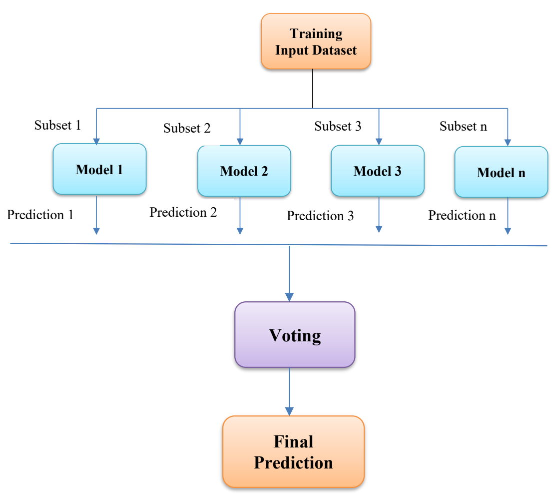

Fig. 4. Flowchart of ensemble bagged tree algorithm.

Overfitting issues can be effectively avoided when this method is used. The complete process of ensemble learning (bagged trees) algorithm is depicted in Fig. 4. As each tree grows independently, training is parallelized and thus the computation speed also accelerates. Bagging is typically robust to noise and outliers because these errors are averaged out. However, due to working of multiple trees, ensemble's logic is hard to deduce unlike single trees.

3. System Modelling

The PV system has been modeled by simulating a nine-panel system arranged in 3 rows and 3 columns. The total power rating of the system is 2kW. Two identical PV system sub-models have been created in the MATLAB-Simulink environment, where the first sub-system represents the physical PV system under non-ideal and shaded conditions in different panels, while the second sub-system represents the ideal, non-shaded system. For increasing the output voltage level of the PV system module, a DC-DC boost converter has been used. To extract the system's maximum power under varying atmospheric conditions, the Perturb and Observe algorithm [45] has been implemented in the model. The schematic diagram of the modeled system is shown in Fig. 5.

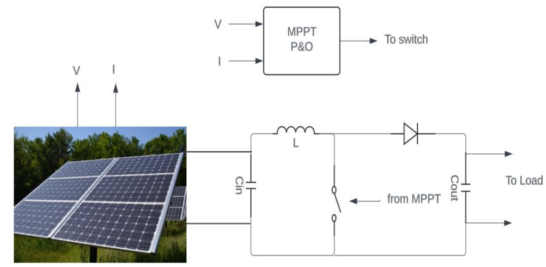

Fig. 5. Schematic diagram of modeled PV system.

The power outputs from each panel of both the sub-modules are measured. The modeled PV system is simulated with a temperature variation from 25°C to 55°C and an irradiance range from 25 to 1000W/m^2. For simplification, the shading scenario is modeled by assuming uniform shading across the entire panel. Table 1 presents the PV system parameters used for system modeling.

Table 1. PV system parameters.

System parameters	Values
Panel voltage at maximum power point (in Volts) for STC (1000 W/m^2 irradiance and 25°C temperature)	29
Open circuit panel voltage (in Volts) for STC	36.3
Panel current at maximum power point (in Amperes) for STC	7.35
Short circuit panel current (in Amperes) for STC	7.84
Diode ideality factor	0.98117
Cells per module	60

3.1. Process 1: Identification of the highest shaded panel

Various shading scenarios can occur in PV panels, ranging from shading of a single panel to multiple panels simultaneously. For maintenance purposes, priority should be given to the panel experiencing the highest shading. The methodology for identifying and addressing such cases is outlined here. The voltage and current outputs of each panel are measured. The product of voltage and current, i.e., the power obtained from each panel, is measured from both the sub-models. For dataset generation, the predictors considered are mentioned in Table 2. To identify the highest shaded panel, simulation has been done by considering three different cases for every individual panel, i.e.,

1) when only one panel is shaded,
2) when two panels are shaded with different shading severity , and
3) when three panels are shaded with different shading severity.

This process can be extended by taking more than three panels as well, but in this paper, we have limited the study by taking the shading effect up to three different panels at different shading severity and identifying the panel with the highest shading. Based on the given predictors, nine labels are defined: Label 1 indicates that the first panel has the highest shading, Label 2 indicates that the second panel has the highest shading, and so on, up to Label

9 for the ninth panel. The details of the dataset are given in Table 3, mentioning the definition of each Label and the dataset size for each panel.

Table 2. Predictors for determining the panel having the highest shading (process 1).

S. No.	Predictors for machine learning algorithm
1	Irradiance
2	Temperature
3	Power outputs of panels 1 to 9 in sub-model 1 subjected to shading effects (9 predictors)
4	Power outputs of panels 1 to 9 in sub-model 2 without shading effects (9 predictors)
5	Relative error of power outputs of sub-model 1 and sub-model 2 (9 predictors)
6	Maximum difference of relative error between sub-model 1 and sub-model 2
7	Lowest recorded power output among panels 1 to 9 subjected to partial shading

Table 3. Dataset label description for process1 (identification of highest shaded panel).

S. No.	Label	Label description	Sample size
1	"1"	Panel 1 is highest shaded	1340
2	"2"	Panel 2 is highest shaded	1400
3	"3"	Panel 3 is highest shaded	1384
4	"4"	Panel 4 is highest shaded	1358
5	"5"	Panel 5 is highest shaded	1316
6	"6"	Panel 6 is highest shaded	1391
7	"7"	Panel 7 is highest shaded	1341
8	"8"	Panel 8 is highest shaded	1382
9	"9"	Panel 9 is highest shaded	1394

3.2. Process 2: Level of shading

It is important to know the extent of shading of PV panels to assess the severity of the situation and plan maintenance accordingly. Therefore, a study was conducted using the same Simulink model, focusing on Panel 1 for analysis. The same methodology can be applied to analyze the extent of shading for all other panels. In this case, the panel is subjected to different shading levels from 0% to 100% by changing the irradiance value. The power output of each panel in sub-model 1 and sub-model 2 is recorded for temperature variations between 15°C and 50°C and irradiance levels ranging from 25 to 1000 W/m^2. The collected data is subsequently used as input for a machine learning algorithm. The predictors used for the machine learning algorithm are mentioned in Table 4. The labeling for shading level identification is defined in Table 5, where each label corresponds to a specific percentage range of shading. Label 1 represents high shading (66–100%), Label 2 represents medium shading (34–65%), and Label 3 represents low shading (0–33%), relative to non-shaded panels.

Table 4. Predictors for determining the shading level of panels.

S. No.	Predictors for machine learning algorithm
1	Irradiance
2	Temperature
3	Power output from panel 1 to 9 in sub-model 1 (with shading effect)
4	Power output from panel 1 to 9 in sub-model 2 (without shading effect)

Table 5. Dataset description for determination of level of shading.

S. No.	Label	Label description	Sample size
1	"1"	High shading (66–100%)	2706
2	"2"	Medium shading (34–65%)	2677
3	"3"	Low shading (0–33%)	2454

Table 6. Sample conditions considered for identification of highest shaded panel.

Shading level of panels	Label
Panel 1 – 90% shaded, other panels – unshaded	1
Panel 1 – 90% shaded, panel 2 – 50% shaded, panel 4 – 20% shaded	1
Panel 1 – 90% shaded, panel 4 – 50% shaded	1
Panel 1 – 90% shaded, panel 4 – 50% shaded, panel 8 – 20% shaded	1
Panel 2 – 90% shaded, other panels – unshaded	2
Panel 2 – 90% shaded, panel 3 – 50% shaded	2
Panel 2 – 90% shaded, panel 4 – 50% shaded	2
Panel 2 – 90% shaded, panel 4 – 50% shaded, panel 7 – 20% shaded	2
Panel 2 – 90% shaded, panel 3 – 50% shaded, panel 4 – 20% shaded	2

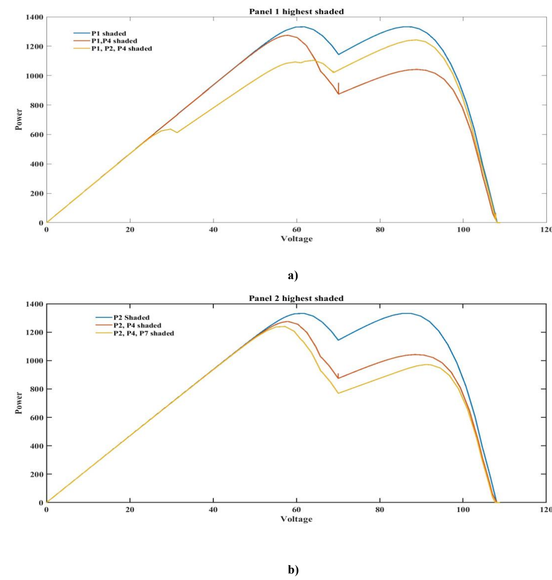

Fig. 6. Shading patterns when multiple panels having unbalanced shading: a) Cases when panel 1 is highest shaded, b) Cases when panel 2 is highest shaded.

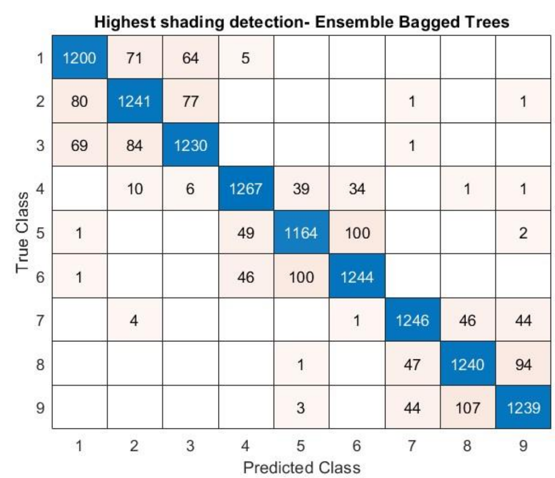

Fig. 7. Confusion matrix for identification of highest shaded panel using ensemble bagged trees algorithm..

3.3. Identification of highest shaded panel

To identify the highest shaded panel, simulation has been carried out in MATLAB-Simulink environment for dataset generation, and the dataset is thereby used for training and testing of different machine learning algorithms. A total of 12,306 samples have been used as datasets for identifying the panel most affected panel by shading. The conditions considered for this study when panels

Table 7. Precision, recall and F1 Score of each label for identification of the highest shaded panel.

Label No.	True positive	False positive	False negative	Precision (%)	Recall (%)	F1 score (%)
1	1200	151	140	88.8	89.6	89.2
2	1241	169	159	88.0	88.6	88.3
3	1230	147	154	89.3	88.9	89.1
4	1267	100	91	92.7	93.3	93.0
5	1164	143	152	89.1	88.4	88.7
6	1244	135	147	90.2	89.4	89.8
7	1246	93	95	93.1	92.9	93.0
8	1240	154	142	89.0	89.7	89.3
9	1239	142	154	89.7	88.9	89.3

Table 8. Performance indices for no, low, medium and high shading levels.

Condition	Rated power (W)	Vm (V)	Im (A)	Voc (V)	Isc (A)	GMPP (W)	% Power output	Mismatch loss (%)	Efficiency (%)
No shading	213.15	89.50	21.184	108.14	23.56	1895.98	98.83	1.16	13.00
Low shading	213.15	85.919	21.660	108.14	23.56	1861.10	97.01	2.98	12.91
Medium shading	213.15	88.366	18.135	108.19	23.56	1602.55	83.53	16.46	11.64
High shading	213.15	85.669	15.551	108.13	23.56	1332.32	69.45	30.54	10.15

Table 9. Precision, recall and F1 score of each label for determination of level of shading.

Label No.	True positive	False positive	False negative	Precision (%)	Recall (%)	F1 score (%)
1	2594	208	114	92.6	95.8	94.2
2	2390	217	290	91.7	89.2	90.4
3	2349	84	106	96.5	95.7	96.1

Table 10. Comparative study: Highest shading detection using different classification learner algorithms.

S. No.	Classification learner	Accuracy (%)	Average precision (%)	Average recall (%)	F1 score (%)
1	Ensemble bagged tree	90.0	89.99	89.97	89.97
2	Ensemble boosted trees	76.8	76.77	78.64	77.70
3	Ensemble subspace discriminant	63.1	63.20	67.61	65.33
4	Ensemble subspace KNN	65.4	65.45	65.40	65.42
5	Fine tree	77.6	77.63	77.92	77.78
6	Quadratic discriminant	84.3	84.45	86.35	85.39
7	Kernel naïve bayes	71.8	71.91	72.79	72.34
8	Quadratic space vector machine	85.9	85.91	86.12	86.01
9	Weighted KNN	59.7	59.73	59.84	59.79
10	Artificial neural network (ANN)	89.6	89.64	89.63	89.64
11	Logistic regression kernel	58.5	58.52	58.44	58.48

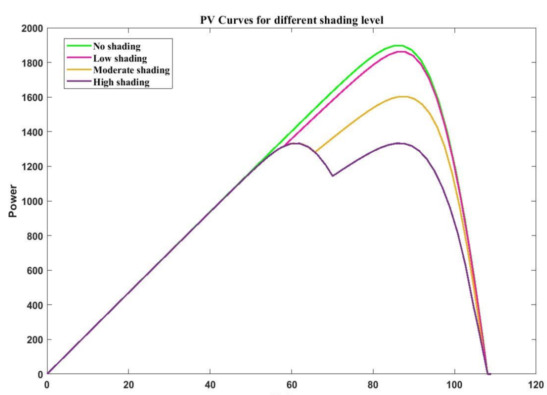

Fig. 8. Shading patterns when a panel receives different level of shading.

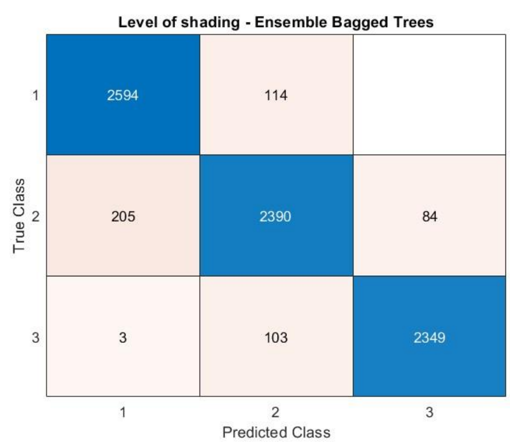

Fig. 9. Confusion matrix for determination of level of shading using ensemble bagged trees algorithm.

1 and 2 have received highest shading has been elaborated in Table 6 where the Labels mentioned have been defined in Table 3. Figs. 6-(a) and 6-(b) show the PV curve patterns for three cases. In Fig. 6-(a): (i) only Panel 1 is shaded, (ii) Panels 1 and 4 are shaded, with Panel 1 having a higher level of partial shading, and (iii) Panels 1 and 4 are shaded, with Panel 1 being the most shaded panel. Similarly, in Fig. 6-(b), three PV curve patterns represent the following cases: (i) only Panel 2 is shaded,

Table 11. Comparative study: level of shading identification using different classification learner algorithms.

S. No.	Classification learner	Accuracy (%)	Average precision (%)	Average recall (%)	F1 score (%)
1	Ensemble bagged tree	93.5	93.60	93.57	93.57
2	Ensemble boosted trees	78.8	78.95	79.27	78.99
3	Ensemble subspace discriminant	61.8	61.69	68.26	62.68
4	Ensemble subspace KNN	79.0	79.15	79.06	79.10
5	Fine tree	86.6	86.68	86.46	86.49
6	Quadratic discriminant	47.5	47.61	65.09	41.88
7	Kernel naïve bayes	51.3	51.95	48.08	42.40
8	Quadratic space vector machine	82.3	82.56	82.53	82.22
9	Weighted KNN	79.1	79.22	79.21	79.21
10	Artificial neural network (ANN)	90.1	90.20	90.23	90.16
11	Logistic regression kernel	73.5	73.73	73.57	73.63

(ii) Panels 2 and 4 are shaded, with Panel 2 having a higher level of partial shading, and (iii) Panels 2, 4, and 7 are shaded, with Panel 2 being the most shaded panel. Similar graphs can be plotted for other panels as well. The PV curves in Fig. 6 illustrate that under non-uniform shading across different panels in a PV array, identifying the most severely affected panel using the PV curve alone is challenging, as the resulting waveforms for such cases are similar in nature. This complexity increases further as the number of shaded panels increases. Thus, for the identification of the most severely panel affected from partial shading when the shading pattern is imbalanced has been found out using ensemble learning algorithm and Fig. 7 shows the confusion matrix obtained while using ensemble bagged trees algorithm's where the classes represent the labels, as mentioned in Table 3.

The results from the confusion matrix shows that the ensemble bagged tree algorithm can identify the highest shaded panel with an accuracy of 90.0%. It can also be observed from the confusion matrix that the mis-classifications in the system are mainly within the panels of the same string.

The performance of the learner is further analyzed based on different performance indices which can be defined as mentioned below:

• Classification Accuracy: It can be defined as the ratio of number of correct predictions to overall number of input samples [46]. It can be calculated using the following expression:

$$\text{Classification Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{True Positives} + \text{False Positives} + \text{False Negatives} + \text{True Negatives}} \tag{1}$$

• Precision: It is the ratio of all the correctly predicted actual true samples to all samples which have been classified as true. It's expression is given as [47]:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{2}$$

• Recall: It is the ratio of all the true positive sample predictions to all the actual positive instances in the total dataset. It can be given as [47]:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{3}$$

• F score: This is the harmonic mean of precision and recall and can be mathematically gives as:

$$\text{F Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

The measured values for the above-said parameters are mentioned in Table 7. The performance of a classification model should be evaluated not only by accuracy but also by precision, recall, and F1 score. Values between 80% and 100% for these metrics indicate excellent performance of the algorithm [48]. Hence it can be observed from Table 7 that the performance matrices of ensemble bagged trees algorithm for identifying the highest shaded panel is lying in the 80-100% range depicting an excellent performance of the algorithm.

3.4. Identification of shading level

In process 2, simulation was done to identify the level of shading for three different categories, which are also termed as low (0-33%) identified as class-3, medium (34-65%) identified as class- 2 and high shading (66-100%) identified as class-1 in the confusion matrix. Fig. 8 shows the shading pattern obtained for panel 1 when encountered by different shading levels i.e. low, medium or high as defined . Similar shading patters can be obtained for other panels as well. The PV curves in Fig. 8 clearly indicate that, compared to normal operation without shading, the power output decreases under shading and declines further to significantly lower values as shading intensity increases. This analysis is crucial for determining maintenance schedules, and prompt action is required for panels exhibiting high shading levels . Fig. 9 shows the output represented by the confusion matrix obtained using ensemble bagged trees algorithm. The accuracy obtained for shading percentage identification through this process is 93.5%.

Furthermore, an analysis has been done regarding the impact of level of shading on the performance of the PV panels by computing the output power, power loss and fill factor for different shading levels. These values can be calculated using the given formulas [21]:

Global maximum power point (GMPP): it is defined as the product of maximum voltage and maximum current for entire PV array.

$$\text{Global maximum power point} = V_m * I_m \tag{5}$$

Output power (P_{out}): It can be defined as the ratio of GMPP to the product of PV panel power rating and number of panels in the array.

$$P_{out}\% = \text{GMPP} * \frac{100}{\text{PV panel power rating} * \text{no. of panels in the array}} \tag{6}$$

Power loss: It is the subtraction of output power in percentage from 100. It indicates the mismatch loss for the array.

$$\text{Power Loss} = 100 - P_{out} \tag{7}$$

Fill Factor (FF): It can be defined as the ratio of product of maximum voltage and maximum current to product of open circuit voltage and short circuit current for the PV array.

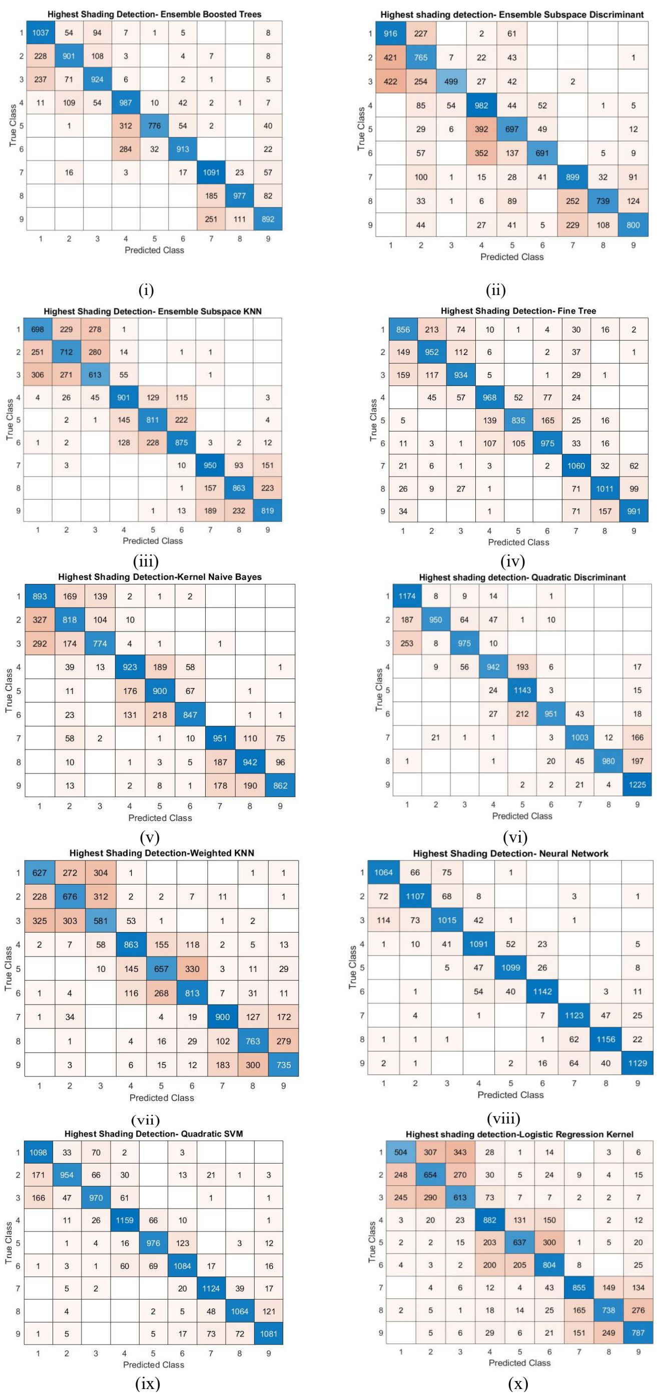

Fig. 10. (i-x) confusion matrices of 10 different classification algorithms for highest shading detection (i) Ensemble boosted trees (ii) Ensemble subspace discriminant (iii) Ensemble subspace KNN (iv) Fine tree (v) Kernel naïve bayes (vi) Quadratic discriminant (vii) Weighted KNN (viii) Neural network (ix) Quadratic SVM (x) Logistic regression kernel.

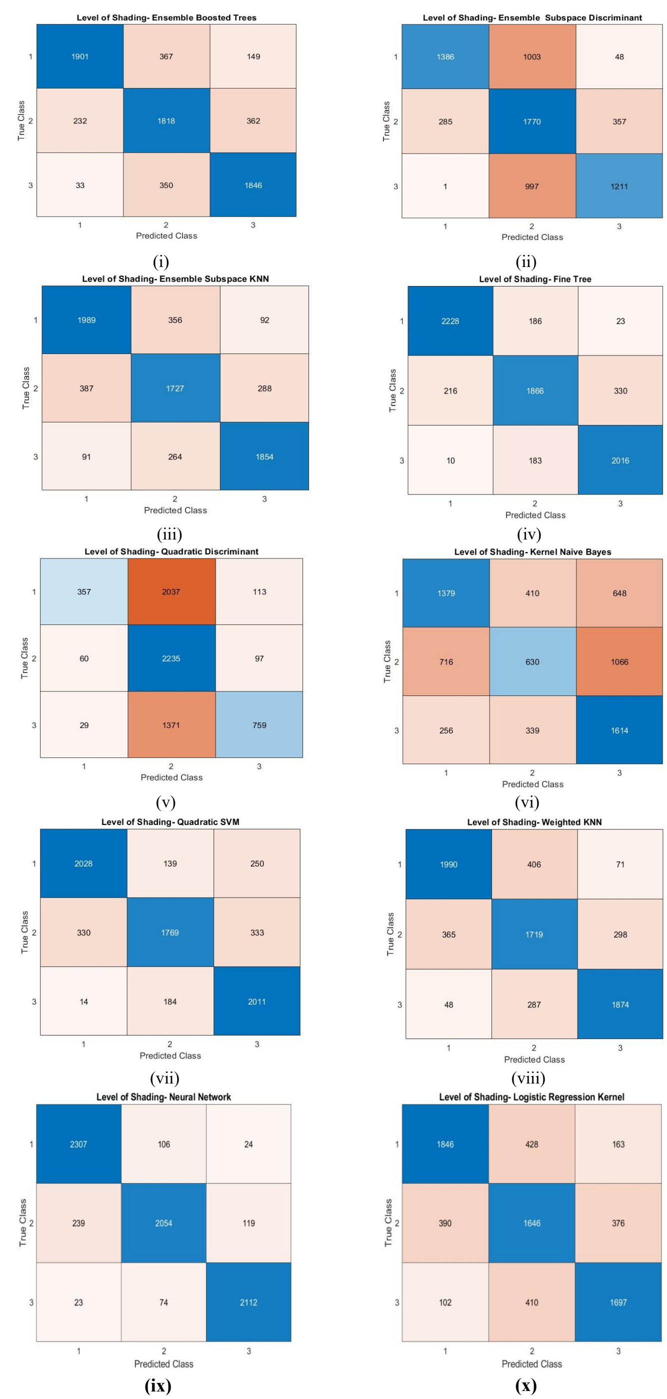

Fig. 11. (i-x) Confusion matrices of 10 different classification algorithms for identification of level of shading (i) Ensemble boosted trees (ii) Ensemble subspace discriminant (iii) Ensemble subspace KNN (iv) Fine tree (v) Kernel naïve bayes (vi) Quadratic discriminant (vii) Weighted KNN (viii) Neural network (ix) Quadratic SVM (x) Logistic regression kernel.

$$\text{Fill Factor (FF)} = \frac{V_m * I_m}{V_{OC} * I_{SC}} \tag{8}$$

Efficiency (η) : It can be calculated as the ratio of GMPP power to product of area of panels A (in m^2) and total irradiance (I_{rt}) received per unit area (in W/m^2).

$$\eta(\text{in}\%) = \frac{\text{GMPP}}{A * I_{rt}} * 100 \tag{9}$$

In this study, the area of a single panel is $1.62m^2$.

Table 8 is generated from simulations where Panel 1 operates either under normal conditions or with varying shading levels i.e., low, medium, or high, as defined in Table 5. The values presented in the table are based on specific shading, irradiance, and temperature conditions. The same approach can be applied to other environmental conditions. It can be deduced from Table 8 that as the shading level increases, the performance of the array deteriorates.

Table 9 mentions the metrics for precision, recall, and F1 score calculated to obtain a greater insight for different labels obtained using the ensemble bagged trees training algorithm. The percent values of precision, recall and F1 score are lying between 80-100% indicating excellent performance of the algorithm for identifying the level of shading.

Furthermore, a comparative study was conducted to evaluate the performance of the ensemble bagged trees algorithm against ten other machine learning/classification algorithms using the same input data. The results, presented in Table 10, show that the ensemble bagged trees algorithm achieved the highest accuracy in identifying both the most shaded panel and the level of shading. The confusion matrices for the ten classification algorithms for identification of highest shaded panel are mentioned in Fig. 10. The accuracy levels, average precision value, recall and F1 score for eleven classification learners are compared and mentioned in Table 10. It can be observed that the values other performance parameters in addition to accuracy, i.e., precision, recall and F1 score is also highest in case of ensemble bagged trees algorithm as compared to other machine learning methods. However, the neural network also demonstrated good performance, achieving an accuracy of 89.6% and precision, recall, and F1 score values of 89.64%, 89.63%, and 89.64%, respectively. A comparative study to identify the level of shading has been carried out for 11 different classification algorithms, including the ensemble bagged trees algorithm, as shown in Table 11. The confusion matrices for these algorithms are depicted in Fig. 11, and the values of accuracy, precision, recall, and F-1 score are calculated through these confusion matrices and have been summarized in Table 11. For identification of the level of shading, the ensemble bagged trees algorithm proves to be superior to the other MLTs with 93.5% accuracy, 93.6% precision value, 93.57% recall value, and an F-1 score of 93.57%.

4. Conclusion

This paper presents a priority-based maintenance strategy for solar panels affected by permanent shading by employing the ensemble bagged trees algorithm for analysis and decision-making. In this paper, the impact of shading on PV panels is illustrated and a method is proposed to identify the highest shaded panel and provide the level of shading of different panels due to permanent shading or soil accumulation. It has been found that among eleven different classification learner algorithms, the ensemble bagged trees algorithm proves to be the most accurate learner. The ensemble bagged trees algorithm can identify the panel having the highest shading with 90.0% accuracy, whereas the accuracy in identifying the level of shading is 93.5%, and the supremacy of the algorithm compared to ten other machine learning algorithms is also demonstrated in the paper. This methodology can greatly improve the performance of the entire system by timely supervision of panels requiring early monitoring and pre-determination of faults due to the shading effect. The precise information about the panels affected by shading, the level of imbalance in terms of shading for different panels, as well as the intensity of shading, is crucial and can be further utilized for fault detection and its early mitigation, as if left unattended, partial shading can deteriorate the performance of the system, reduce the power output, and lead to the generation of hotspots. Thus, this ideology can be implemented by making a simulation model of an ideal system and comparing it with the physical PV system. While the proposed methodology is effective for identification and localization of shading effects, the challenge remains in real-time implementation with dynamic environmental scenarios. Also, as the mentioned work proposes that power sensors should be deployed across each panel, this can substantially increase the capital expenditure of the system. Nevertheless, over time, this expense might be compensated for by lower maintenance costs, longer system lifespan, early fault detection, and decreased energy losses. To calculate the return on investment and support the viability of this strategy, a thorough techno-economic analysis is necessary. Future research could explore the integration of forecasting models that utilize environmental conditions and solar irradiance data to predict potential shading patterns in advance. Additionally, other fault scenarios may be included in the system to investigate the applicability of the mentioned technique for precise identification and localization of faulty panels.

COMPETING INTERESTS

The authors affirm that there are no conflicts of interest associated with this manuscript.

FUNDING

The authors declare that no funding is provided to carry out this research work.

References

[1] S. Behzadi, N. Osali, A. Younesi, and A. Bagheri, “A cost-effective trade-off in distribution system expansion planning between construction of conventional/renewable distributed energy sources in long term,” *J. Oper. Autom. Power Eng.*, vol. 13, no. 3, pp. 196–205, 2025.

[2] Energy World, “India now has 70,000 mw of solar power generation capacity; rajasthan leading the pack.” https://energy.economictimes.indiatimes.com/news/renewable/india-now-has-70000-mw-of-solar-power-generation-capacity-rajasthan-leading-the-pack/102577162, 2023. Accessed: 2025-09-01.

[3] M. Dhimish and A. M. Tyrrell, “Photovoltaic bypass diode fault detection using artificial neural networks,” *IEEE Trans. Instrum. Meas.*, vol. 72, 2023.

[4] IEA, “Snapshot 2021.” https://iea-pvps.org/snapshot-reports/snapshot-2021. Online report.

[5] M. S. Syed, C. V. Suresh, and S. Sivanagaraju, “Impact of renewable sources on electrical power system,” *J. Oper. Autom. Power Eng.*, vol. 12, no. 3, pp. 261–268, 2024.

[6] M. Gupta, “Explained: Mitigation techniques for dust and soil effects on solar panel performance.” https://solarquarter.com/2023/05/26/explained-mitigation-techniques-for-dust-and-soil-effects-on-solar-panel-performance/, 2023. Solarquarter.

[7] R. H. Fonseca Alves, G. A. de Deus Júnior, E. G. Marra, and R. P. Lemos, “Automatic fault classification in photovoltaic modules using convolutional neural networks,” *Renew. Energy*, vol. 179, pp. 502–516, 2021.

[8] G. Abdelaziz, H. Hichem, B. R. Chiheb, and G. Rached, “Shading effect on the performance of a photovoltaic panel,” in *2021 IEEE 2nd Int. Conf. Signal, Control Commun.*, pp. 208–213, IEEE, 2021.

[9] N. Khadka, A. Bista, B. Adhikari, A. Shrestha, D. Bista, and B. Adhikary, “Current practices of solar photovoltaic panel cleaning system and future prospects of machine learning implementation,” *IEEE Access*, vol. 8, pp. 135948–135962, 2020.

[10] L. B. Bosman, W. D. Leon-Salas, W. Hutzel, and E. A. Soto, “Pv system predictive maintenance: Challenges, current approaches, and opportunities,” *Energies*, vol. 13, no. 6, p. 1398, 2020.

[11] K. Abdulmawjood, S. Alsadi, S. S. Refaat, and W. G. Morsi, “Characteristic study of solar photovoltaic array under different partial shading conditions,” *IEEE Access*, vol. 10, pp. 6856–6866, 2022.

[12] F. Ramezani and M. Mirhosseini, “Shading impact modeling on photovoltaic panel performance,” *Renew. Sustain. Energy Rev.*, vol. 212, p. 115432, 2025.

[13] M. Khan, M. Ferdous, and S. Chowdhury, "Sustainable development of solar power through the investigation of partial shading effect of solar module in terms of experimental set up and matlab simulation," *Procedia Comput. Sci.*, vol. 252, pp. 702–707, 2025.

[14] F. S. M. Abdallah, M. N. Abdullah, I. Musirin, and A. M. Elshamy, "Intelligent solar panel monitoring system and shading detection using artificial neural networks," *Energy Rep.*, vol. 9, pp. 324–334, 2023.

[15] A. K. Tripathi, M. Aruna, and C. S. N. Murthy, "Performance of a pv panel under different shading strengths," *Int. J. Ambient Energy*, vol. 40, no. 3, pp. 248–253, 2019.

[16] S. Ansari, A. Ayob, M. S. H. Lipu, M. H. M. Saad, and A. Hussain, "A review of monitoring technologies for solar pv systems using data processing modules and transmission protocols: Progress, challenges and prospects," *Sustainability*, vol. 13, no. 15, 2021.

[17] T. Kappler, A. S. Starosta, N. Munzke, B. Schwarz, and M. Hiller, "Detection of shading for solar power forecasting using machine learning techniques," in *Proc. Eur. PV Solar Energy Conf. Exhib.*, 2023.

[18] E. A. Setiawan, M. Fathurrahman, R. F. Pamungkas, and S. Ma'Arif, "Fast partial shading detection on pv modules for precise power loss ratio estimation using digital image processing," *J. Electr. Comput. Eng.*, vol. 2024, p. 9385602, 2024.

[19] F. Belhachat and C. Larbes, "Photovoltaic array reconfiguration strategies for mitigating partial shading effects: Recent advances and perspectives," *Energy Convers. Manage.*, vol. 313, p. 118547, 2024.

[20] S. Ganesan, P. W. David, P. K. Balachandran, and I. Colak, "Power enhancement in pv arrays under partial shaded conditions with different array configuration," *Heliyon*, vol. 10, no. 2, 2024.

[21] R. Verma, S. Gupta, and A. Yadav, "Performance evaluation of different photovoltaic array configurations under partial shading," *Renew. Energy*, vol. 237, p. 121796, 2024.

[22] S. Saravanan, R. S. Kumar, and P. Balakumar, "Binary firefly algorithm based reconfiguration for maximum power extraction under partial shading and machine learning approach for fault detection in solar pv arrays," *Appl. Soft Comput.*, vol. 154, p. 111318, 2024.

[23] Y. Wang and B. Yang, "Optimal pv array reconfiguration under partial shading condition through dynamic leader based collective intelligence," *Prot. Control Mod. Power Syst.*, vol. 8, no. 1, 2023.

[24] V. Jain, R. Singh, R. Yadav, V. K. Yadav, V. Kumar, and S. Garg, "Multi-step optimization for reconfiguration of solar pv array for optimal shade dispersion," *Electr. Eng.*, 2024.

[25] P. R. Satpathy, B. Aljafari, S. B. Thanikanti, N. Nwulu, and R. Sharma, "A multi-string differential power processing based voltage equalizer for partial shading detection and mitigation in pv arrays," *Alex. Eng. J.*, vol. 104, pp. 12–30, 2024.

[26] B. Liu *et al.*, "A novel data-driven state evaluation approach for photovoltaic arrays in uncertain shading scenarios," *Energy*, vol. 312, p. 133533, 2024.

[27] V. Sarkar, V. K. Kolakaluri, and S. Anantha, "Enhancing the maximum or flexible power point tracking control of a photovoltaic array with a non-invasive and computationally robust model-based method for partial shading detection," *Electr. Power Syst. Res.*, vol. 238, p. 111096, 2025.

[28] U. Kumar, S. Mishra, and K. Dash, "An iot and semi-supervised learning-based sensorless technique for panel level solar photovoltaic array fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 72, 2023.

[29] J. Zhang *et al.*, "An i–v characteristic reconstruction-based partial shading diagnosis and quantitative evaluation for photovoltaic strings," *Energy*, vol. 300, p. 131480, 2024.

[30] A. Ahmadpour, A. Dejamkhooy, and H. Shayeghi, "Optimization and modelling of linear fresnel reflector solar concentrator using various methods based on monte carlo ray–trace," *Sol. Energy*, vol. 245, pp. 67–79, 2022.

[31] B. Nomeir, S. Lakhouil, S. Boukheir, M. A. Ali, and S. Naamane, "Recent progress on transparent and self-cleaning surfaces by superhydrophobic coatings deposition to optimize the cleaning process of solar panels," *Sol. Energy Mater. Sol. Cells*, vol. 257, p. 112347, 2023.

[32] Y. Wu *et al.*, "A review of self-cleaning technology to reduce dust and ice accumulation in photovoltaic power generation using superhydrophobic coating," *Renew. Energy*, vol. 185, pp. 1034–1061, 2022.

[33] V. P. Madhanmohan, M. Nandakumar, and A. Saleem, "Enhanced performance of partially shaded photovoltaic arrays using diagonally dispersed total cross tied configuration," *Energy Sources A: Recovery Util. Environ. Eff.*, 2020.

[34] D. Craciunescu and L. Fara, "Investigation of the partial shading effect of photovoltaic panels and optimization of their performance based on high-efficiency flc algorithm," *Energies*, vol. 16, no. 3, 2023.

[35] S. Selvi, M. Mohanraj, P. Duraipandy, S. Kaliappan, L. Natrayan, and N. Vinayagam, "Optimization of solar panel orientation for maximum energy efficiency," in *2023 4th Int. Conf. Smart Electron. Commun.*, pp. 159–162, 2023.

[36] A. K. Gupta, T. Maity, A. H, and Y. K. Chauhan, "An electromagnetic strategy to improve the performance of pv panel under partial shading," *Comput. Electr. Eng.*, vol. 90, p. 106896, 2021.

[37] V. M. R. Tatabhatla, A. Agarwal, and T. Kanumuri, "A chaos map based reconfiguration of solar array to mitigate the effects of partial shading," *IEEE Trans. Energy Convers.*, vol. 37, pp. 811–823, 2022.

[38] H. Oufettoul, N. Lamdihine, S. Motahhir, N. Lamrini, I. A. Abdelmoula, and G. Aniba, "Comparative performance analysis of pv module positions in a solar pv array under partial shading conditions," *IEEE Access*, vol. 11, pp. 12176–12194, 2023.

[39] M. Baghoolizadeh, A. A. Nadooshan, A. Raisi, and E. H. Malekshah, "The effect of photovoltaic shading with ideal tilt angle on the energy cost optimization of a building model in european cities," *Energy Sustainable Dev.*, vol. 71, pp. 505–516, 2022.

[40] L. El Iysaouy *et al.*, "Performance enhancements and modelling of photovoltaic panel configurations during partial shading conditions," *Energy Syst.*, 2023.

[41] M. A. Raza *et al.*, "Mitigating the impact of partial shading conditions on photovoltaic arrays through modified bridge-linked configuration," *Sustainability*, vol. 17, no. 3, 2025.

[42] P. Venkata, V. Pandya, and A. V. Sant, "Data mining and svm based fault diagnostic analysis in modern power system using time and frequency series parameters calculated from full-cycle moving window," *J. Oper. Autom. Power Eng.*, vol. 12, no. 3, pp. 206–214, 2024.

[43] V. Singh, A. Yadav, S. Gupta, and A. Y. Abdelaziz, "Switch fault identification scheme based on machine learning algorithms for pv-fed three-phase neutral point clamped inverter," *e-Prime Adv. Electr. Eng. Electron. Energy*, vol. 8, 2024.

[44] A. Nagpal, "Decision tree ensembles- bagging and boosting." https://medium.com/towards-data-science/decision-tree-ensembles-bagging-and-boosting-266a8ba60fd9, 2020.

[45] A. Raj and R. P. Praveen, "Highly efficient dc-dc boost converter implemented with improved mppt algorithm for utility level photovoltaic applications," *Ain Shams Eng. J.*, vol. 13, no. 3, 2022.

[46] S. M. Pizer and J. S. Marron, "Object statistics on curved manifolds," in *Stat. Shape Deform. Anal.: Methods*

Implementation Appl., pp. 137–164, 2017.
[47] N. Acharya, "Understanding precision, recall, f1-score, and support in machine learning evaluation." https://medium.com/@nirajan.acharya777/understanding-precision-recall-f1-score-and-support-in-machine-learning-evaluation-7ec935e8512e, 2023.
[48] Keylabs, "Understanding the f1 score and auc-roc curve." https://keylabs.ai/blog/understanding-the-f1-score-and-auc-roc-curve/, 2023.