Engineering Systems and Intelligent Technologies
ESIT

Journal Homepage
https://pub.scientificirg.com/index.php/ESIT

CVaR-Optimized Distributionally Robust Stacked Ensemble With Range-Based Current Signatures for Reliable Fault Detection in Photovoltaic Farms

Yasser Abdel Satar[a,1] , **Hadeel A. S. Almansouri** [b], **Amr A. Hassanain**[a]

[a] *Department of Artificial Intelligence, Faculty of Computers and Artificial Intelligence Sphinx University, Assiut 71511, Egypt.*
E-mail: yasser.selim@sphinx.edu.eg, Amr.Hassanain@sphinx.edu.eg.
[b] *College of Computers and Information Technology, Taif University, Saudi Arabia. E-mail: Hadeel.mns97@gmail.com.*

ABSTRACT

Reliable photovoltaic (PV) Fault Detection Prevention of energy loss, safety, cost reduction. Current methods lack analysis of PV characteristics and operational variability. Remote monitoring does not consider PV farm current. This study develops a new machine learning framework for strain fault diagnosis. Analyzes patterns in diagnoses PV farm central current and string current. Uses current range imbalance techniques. Stacked heterogeneous machine learning uses Extra Trees, boosting, and support vector machines. Uses probability-level fusion. Other techniques used could enhance decision boundaries. Operating imbalance used a loss approach for fault detection. Focuses on worst-case scenarios without operational restrictions. Evaluated on a simulated-data PV plant of 250 kW with variable parameters of irradiance, temperature, and fault resistance. CVaR provides better results compared to common base methods and achieves AROC of 99.3. Maintains. 98.3 specific false alarm rate. Flexible approach to operational variability. Machine learning framework provides multiple string fault PV diagnosis. Combining physically informed feature engineering, stacked ensemble learning, and robustness-aware tuning has proven effective and practical for improving fault-detection reliability in grid-connected PV systems.

PAPER INFORMATION

HISTORY
Received: 25 August 2025
Revised: 15 November 2025
Accepted: 6 February 2026
Online: 26 February 2026

MSC
62K05; 62K15

KEYWORDS
Photovoltaic farm;
Fault detection;
Ensemble Learning;
Distributionally Robust Optimization;
Fault classification.

1 Introduction

From the start to the end of 2023, the cumulative global expansion of the photovoltaic (PV) generation exceeding 1.6 terawatts has to a large part altered the world's energy landscape [1]. As utility-scale PV farms are built globally, the need for these installations to operate without faults is key to their contribution to the stability of the grid. However, these installations face a number of persistent operational reliability challenges. Electrical faults can not only pose a fire safety

[1]Department of Artificial Intelligence, Faculty of Computers and Artificial Intelligence Sphinx University, Assiut 71511, Egypt.
E-mail: yasser.selim@sphinx.edu.eg.

risk but can also lead to 15-30% reductions in energy yield annually [2]. For these reasons, automated fault detection is critical for both the operational safety and the economic viability [3]. The CVaR-optimized stacked ensemble achieves exemplary performance metrics, particularly an AROC of 0.993 and a specificity of 0.983, making it an exemplar in the current data-driven PV fault detection domain. Most recent studies using Convolutional Neural Networks (CNNs) for defect identification from infrared images report outstanding performance metrics, though they are limited by the reliance on expensive thermal cameras and are not ideal for ongoing electrical monitoring [4, 5]. Likewise, hybrid approaches that integrate CNNs with optimization techniques, such as Bitterling Fish Optimization (BFO), achieve high classification scores on image datasets but, critically, fail to account for the robustness of the model to distributional shifts expected in operational PV farms [6]. On the other end of the spectrum, methods using electrical time-series data, such as comparative studies of ensemble methods, demonstrate the operational effectiveness of XGBoost, but typically set optimization goals for average performance on fixed datasets, thereby leaving unassessed how reliable the performance is under changing environmental conditions [7, 8]. This work makes an important contribution by not only achieving the high discriminative accuracy of current state-of-the-art methods - using only basic electrical measurements - but also incorporating a distributionally robust optimization (DRO) approach.

Concerning the tuning procedure in which Conditional Value-at-Risk (CVaR) is minimized, our model shows optimization of the worst case scenario, addressing perhaps the most prominent critique of the field deployment of machine learning for renewable energy systems [9]. Thus, the tool we design meets a cutting edge standard in detection, while the model design is for the most severe real world variab le operating conditions.

Conventional methods of PV monitoring have a number of limitations. Large scale installations tend to be economically impractical with manual inspections being very comprehensive [10]. Model based methods suffer from parameter ambiguity and model analytical environmental variability [11]. Processing signals in spectroscopy and impedance requires advanced equipment with high sample rates which limits scale to [12]. These constraints have inspired alternative utilizing machine learning automated diagnosis from standard electrical measurements.

In the past few years the advancements in PV fault detection that is data driven have grown considerably. In the earlier days of machine learning, there was the use of traditional methods such as Support Vector Machines (SVM) [13], decision trees [14], and Artificial Neural Networks (ANN) [15]. Ensemble methods showed better results and was particularly true in the use of random forests and gradient boosting, as they were able to detect and capture the non-linear electrical data [16]. Other methods built on the use of Deep learning architectures, with the use of Convolutional networks in recognizing spatial patterns and the use of recurrent networks in the capture of temporal dependencies [17]. As shown in **Figure 1** the engineering of features has developed, with the current studies pointing out the statistical current signatures for the detection faults at the string level.

Figure 1: Schematic/Context of a grid-connected utility-scale PV farm.

Advancements made do not fill the important gaps in practical deployment. The first of such gaps is that most existing models optimize for average performance while ignoring distributional shifts brought about by seasonal changes, sensor deterioration, or changes in the configuration of the arrays . This leads to operational-drift-induced performance loss that further leads to a compromise in reliability. Also, there is a majority of studies that center attention on simplified binary detection of faults and do not delve into multi-class discrimination of common faults and differing degrees of severity.

Third, real-time monitoring systems run into computational bottlenecks from complex feature engineering . Finally, model developments that take into account the need for field stability and performance gaps are few and far in between, and this need is most evident [9].

This study addresses these limitations through three integrated contributions:

1. A compact feature set centered on range-based current signatures that capture string-level imbalance with minimal computational overhead.
2. A heterogeneous stacked ensemble combining Extra Trees, histogram-based gradient boosting, and RBF-SVM through probability-level fusion for complementary decision boundaries.
3. A DRO framework using Conditional Value-at-Risk (CVaR) to prioritize worst-case performance under plausible operating shifts.

The study focuses on diagnosing string-level faults (e.g., F1, F2, F3) within such arrays using electrical measurements from the combiner box and inverter DC side. In order to achieve both detection precision and operational consistency for the multi-class fault discrimination problem at utility-scale PV farms, the framework is evaluated on a simulated 250-kW plant dataset with varying irradiance, temperature, and fault resistance conditions.

2 Related Work

Detecting issues in PV modules has advanced as solar installations have become more intricate and large in scale. Initially, manual checks as well as basic electrical inspections were automated based on new advancements in computer technology and accessible data [10]. In this section, I will analyze the changes from model-based systems to the current data-driven systems that incorporate traditional machine learning, deep learning, and other forms of machine learning. Most studies have looked at the data-driven systems, however have not looked as much at the systems that are model based. Most studies have not looked as much at model-based systems. Most existing literature examines data-driven systems, leaving model-based systems understudied. This leaves a large gap in much of the relevant literature. This review's focus is on addressing this gap utilizing, as a primary focus, the integrating of principles of distributionally robust optimization into a stacked ensemble machine learning model for the first time in this field.

2.1 From Physical Models to Data-Driven Paradigms

Research into automated PV fault diagnosis has been **model-based approaches**. These approaches utilize mathematical PV system models, one being the single-diode model – and the other the double-diode model – and simulate expected system behavior for given environment scenarios. Simulation outputs are compared with measured parameters like current, voltage, and power, and fault identifications are made from large residual deviations [11]. While these methods are interpretable, they are also very limited. First, they are most dependent on the accuracy of model parameters, which are likely to be affected by aging and the manufacturing of model dvt [18]. Second, given the same parameters for the model, they may not be able to differentiate between a short-circuit fault of some moderate resistance and partial shading, and with dynamically changing boundary conditions of irradiance and temperature they may produce missed detections or false alarms .

The advancement of supervisory monitoring systems made it possible to shift to **data-driven techniques**, which learn the patterns of normal and faulty operation from historical or simulated data and do not require an analytical physical model. This approach can be further divided into a number of developing branches.

2.1.1 Conventional Machine Learning and Feature Engineering

A large volume of research has utilized supervised machine learning (ML) algorithms for fault classification using processed features from specialized electrical time-series data. Some example features include processed time series data, statistical moments (mean, variance, and skewness), ratios of current/voltage, and power loss metrics derived from string or array level measurements. SVM, k-Nearest Neighbors (kNN), Decision Trees (DT), and ensemble methods such as Random Forest (RF) have been widely researched [13]. A recent study is, for example, a recent study that analyzed ensemble methods such as XGBoost regarding their ability to detect line-line and ground faults, and it was one of the most accurate classifiers [7]. These methods have the advantage of low computational cost and high interpretability while the main barrier is the quality of the engineered features. This is also where the main challenge lies as more often than not, these features require a robust domain expertise to create and because of this, the created features often fail to generalize to a given fault and/or operational scenario .

2.1.2 Deep Learning and Image-Based Diagnosis

Deep Learning (DL) is used in some research, specifically CNNs, to perform fault diagnosis from the images it receives. The dominant input types of the images captured are the infrared (IR) thermography and electroluminescence (EL) imaging. Advanced imaging techniques capture the heat from cracks in the cells and the faults of bypass diodes and extreme mismatches [4]. The recent research is gaining the attention of the community because of the unique challenges posed by imaging the thermals, such as low contrast, great intra-class resemblance, and extreme imbalance of the classes. Two-stage frameworks (the first stage as an anomaly detector, while the second as an anomaly type classifier) coupled with Class Imbalance Learning (CIL) are some of the examples of advanced frameworks [5]. In their attempts to reduce the class imbalance, the majority of these vision-based frameworks achieve great diagnostic accuracy, however they tend to rely on high end (and expensive) technology (IE. drones equipped with infrared cameras) and are not appropriate for an environment where real-time monitoring is conducted using the electrical measurements on a continuous basis [3].

2.1.3 Signal Processing and Hybrid Methods

In an effort to employ the potential of DL to multiple sets of standard electrical data, researchers designed techniques to convert one-dimensional (1D) successive current and voltage signal data into two-dimensional (2D) image-like data. Time-frequency scalograms generated through Continuous Wavelet Transform (CWT) are used as inputs to various CNN architectures for automatic feature extraction and data classification [17]. This successfully foregrounds and encapsulates some of the temporal and spectral fault signatures. In a bid to address the computational constraints for edge deployment, some recent works propose lightweight CNNs, the architectures of which are optimized using metaheuristic approaches . For some of the deep feature extractor hybrid methods with nature inspired optimization for feature selection, a singular high accuracy for a particular defect classification has been reported [6]. While the deep learning and signal processing methods described above are quite advanced, they often tend to be single model systems, with their ease of use and operational approach not being a particular design objective.

2.2 The Pursuit of Robustness and the Ensemble Advantage

Two interconnected challenges emerge from the literature, framing the core contribution of this work.

1. The Distributive Robustness Gaps: The leading models, ML or DL, are trained and tested on fixed data sets and implicitly take as given that the training and operational datasets are drawn from the same distribution. This assumption often does not hold on PV farms due to the seasonal and diurnal cycles, sensor miscalibration, gradual module degradation, and the emergence of novel noise patterns. A model that excels in average accuracy may experience substantial decreases in performance when operating in ever-shifting environments, leading to unreliable predictions [9]. Although DRO is a well-known concept in the broader planning of energy systems, for example, the use of Wasserstein ambiguity sets or Conditional Value-at-Risk (CVaR) to account for uncertainty in the integration of renewables , there are still gaps in its constructive application in the model selection and hyperparameter tuning of PV fault classifiers.

2. Ensemble Learning as a Pathway to Stability: Methods to ensemble, such as stacking, are the combining of predictions of several different base learners (e.g. tree-based models, kernel methods) and building a meta model. This construction, by virtue of leveraging the individual models' different strengths, can to reduce variance, overfitting and achieve greater accuracy than any one individual model constituent. This is the case [16] and the principal reason most practitioners will build models and then optimize those ensembles for average CV (e.g. mean F1 score) performance. There is then a novel opportunity to build the ensemble construction and tuning based on a robustness-oriented objective function. More precisely, the model constructed to minimize CVaR of the loss over the set of validation scenarios balanced for the most likely operational perturbations should yield a model for which accuracy on average is eclipsed by reliability during adverse conditions.

2.3 Comparative Synthesis of Recent Advances

The subsequent table summarizes and contrasts ten central studies from 2023-2026, pinpointing and tracking the shifts in methodology and the gradual evolution toward the robustness-centered approach advocated in this paper.

Table1 shows comparative analysis of recent PV studies. The domain is moving from prioritizing achieving maximum accuracy in a controlled environment to solving practical problems such as computational [19], class imbalance [5], and operational electrical data [7]. Nevertheless, an important and unique gap remains at the crossroads of high-performance electrical data classification, ensemble learning, and explicit distributional robustness. There is a gap in the literature which systematically reframes the problem of hyperparameter tuning and model selection for an ensemble of PV fault detection models under a DRO-inspired objective such as CVaR. This is concerning as the models may suffer from performance

deficiency in exactly the operational conditions they are meant to endure. We suggest using a stacked ensemble classifier, which is built on discriminative, range-based electrical current signatures (a low-cost and always accessible data resource). The most important part of this work is the implementation of a CVaR-based robust optimization step directly into the ensemble tuning procedure. This makes certain that the chosen model is tailored to the best performance on most of the metrics, but also worst-case scenarios on the concerns of the operational variability that are encountered in the real world. This balances the methodological rigor of DRO from energy systems planning [20] with the meticulousness needed for dependable fault diagnostics at the string level, and paves the way to more resilient and reliable PV monitoring systems.

Table 1: Comparative Analysis of Recent PV Fault Detection Studies (2023–2026)

Study (Author, Year)	Core Methodology	Key Strengths	Notable Limitations / Research Gap Highlighted
Sabati et al. (2025) [6]	CNN (VGGNet-16) with Bitterling Fish Optimization (BFO) for feature selection.	High accuracy (98.75%) reported for panel defect classification; effective dimensionality reduction.	Focused on image-based defects; robustness under varying electrical conditions not explored.
Bougoffa M et al., (2025) [19]	Lightweight CNN architecture optimized by Energy Valley Optimizer (EVO) for CWT scalograms.	Designed for efficient edge deployment; uses rich time-frequency features from 1D signals.	Architecture is a single, optimized model. Does not employ ensemble diversity or robust optimization.
Hamad & Sheikholeslami (2024) [7]	Comparative study of supervised ML (XGBoost, SVM, RF, DT) on electrical time-series data.	Strong benchmarking of interpretable models; highlights effectiveness of ensemble methods like XGBoost.	Focuses on average performance on a static dataset; does not address robustness to distribution shift.
Arenas-Prado et al. (2025) [5]	Hierarchical deep learning framework with contrast-preserving data augmentation for imbalanced IR datasets.	Directly addresses critical class imbalance problem in thermal diagnostics; improves minority class recognition.	Inherently tied to thermal image input; computational focus is on vision, not electrical signature robustness.
Castellanos et al. (2025) [20]	DRO for microgrid energy management and tariff design.	Provides a formal mathematical framework (e.g., CVaR) for decision-making under uncertainty in energy systems.	Applied at the system planning/economic level, not at the component-level fault classifier tuning stage.
Samanta et al. (2023) [17]	1D-CNN and LSTM networks for fault diagnosis using raw electrical time-series data.	Captures temporal dependencies in fault progression; end-to-end learning from raw data.	Model performance and generalization can be sensitive to hyperparameters and data quality shifts.
Wang et al. (2025) [9]	Review and analysis of robustness challenges in machine learning for renewable energy applications.	Clearly articulates the problem of distributional shift and its impact on model reliability in field deployment.	A review paper that identifies the gap but does not propose a specific methodological solution.
Haidari et al. (2022) [4]	Deep learning (EfficientNet feature extraction + SVM) for classification of infrared thermal images.	Effective for visual fault classification; demonstrates transfer learning for thermal data.	Relies on costly thermal imaging; not suitable for continuous, SCADA-based electrical monitoring.
Dhimish et al. (2021) [10]	Comprehensive review of catastrophic faults in PV arrays, including detection and mitigation techniques.	Broad overview of fault types, causes, and the landscape of detection technologies.	Scope is general, covering all techniques, and thus does not delve deep into algorithmic robustness.
Zwirtes et al. (2025) [13]	Machine learning for PV fault detection: A comprehensive review of algorithms and feature sets.	Detailed survey of conventional ML applications and feature engineering strategies in the field.	Predates the latest advances in deep learning and robust optimization for this specific problem.

3 Methodology

This part outlines the complete method that was created for the efficient diagnosis of faults in PV strings. **Figure 2** shows the integrated pipeline's high-level overview which spans from data generation and feature engineering to the building of a robust stacked ensemble model and ends with a thorough evaluation. The methodology's focus is on the primary issue of dependable fault classification discernment for differing operational states that characterize the PV farm setting. The pipeline encompasses data simulation, feature engineering, model development with DRO, and multi-faceted performance evaluation.

3.1 Research Design and Experimental Framework

The study participants conduct an experiment with an data-driven methodology with an supervised learning approach for multi-class fault diagnosis. A simulated model of a 250-kW grid connected PV plant will be the testbed. It will be able to generate time-series data that will be consistent with the controlled physically with the strings and electric fault modes at normal operations and at three different levels. The ability to simulate these environments will help in the the control of the in the specifics of the fault attributes of environmental (focal, temperature) and fault (focal, location, variable) parameters and will help to in develop a data set that is thoroughly labeled for the model. [11]

To guarantee methodical and reproducible results, the experimental flow has been divided into different sequential steps:

1. **Data Generation and Partitioning:** To avoid data leakage, independent training and testing datasets are created with respect to simulated fault inception time using non-overlapping time windows.
2. **Preprocessing and Feature Engineering:** Signals of the raw current and voltage are processed and from the range-based signature which detects level-synchronized current imbalance, a compact and discriminative feature set is created.

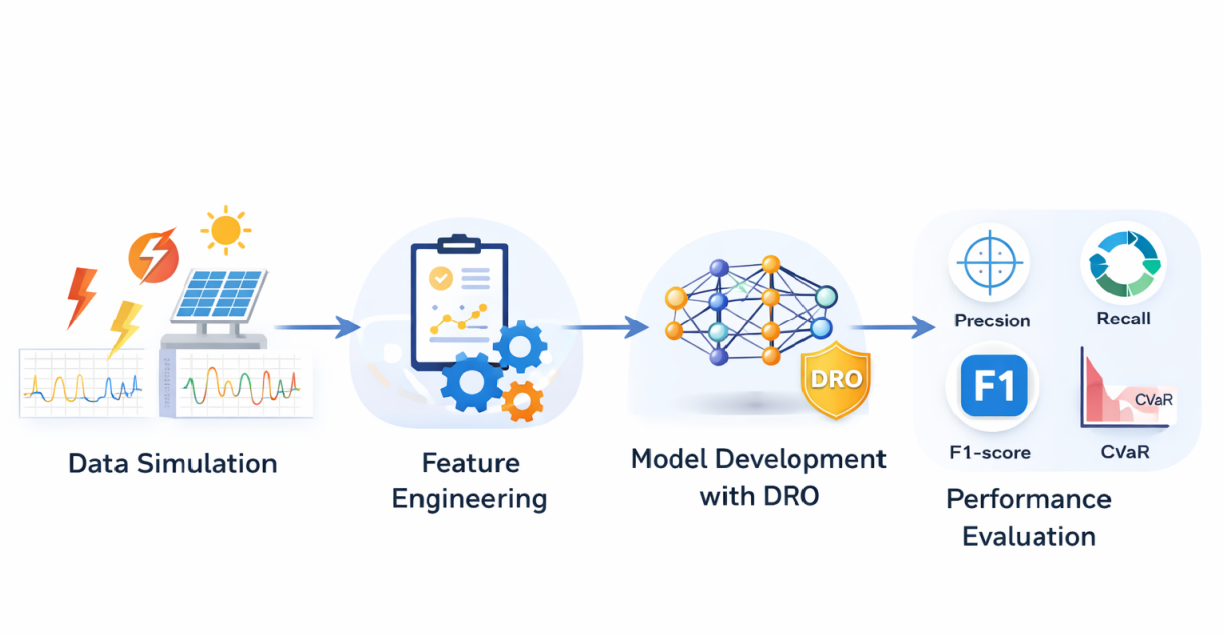

Figure 2: Overview of the proposed methodology for CVaR-optimized PV fault diagnosis.

3. **Baseline Model Establishment:** To set the performance baselines, standard machine learning classifiers are trained and evaluated using the same data and evaluation protocol.
4. **Stacked Ensemble Development:** Using probability-level meta-learning for base learners fusion, predictive accuracy and consistency are improved and integrated into a heterogeneous stacked ensemble.
5. **Robust Hyperparameter Optimization:** For the ensemble based on the CVaR (Conditional Value-At-Risk) of the DRO strategy, worst-case validation scenarios are tailored to prioritize reliable performance during uncertain real-world operational shifts.
6. **Performance Evaluation and Analysis:** For model selection, the generalization is checked using stratified 5-fold cross-validation, followed by a review on the metrics using a completely isolated test set to measure the class-balanced performance and overall discrimination from a range of presets.

   The main goal of this study is to determine empirically if the fault discrimination of the proposed stacked ensemble with CVaR-based robust tuning is both more reliable and superior compared to individual baseline models and a standard-tuned ensemble.

3.2 Dataset Description and Feature Engineering

Adetailed simulation of a 250-kW PV power plant, configured with multiple parallel strings, was conducted to generate the fault diagnosis dataset. Four distinct operational states were defined: one fault-free (normal) class and three string-level fault classes. The class distribution within the training set is detailed in **Table 2**.

Table 2: Class distribution in the training dataset (600 instances).

Description	Class Label	Count	Percentage
Fault-free (Normal) Operation	0	100	16.67%
String Fault (F1)	1	153	25.50%
String-to-Ground Fault (F2)	2	149	24.83%
String-to-String Fault (F3)	3	198	33.00%
Total		**600**	**100%**

The unprocessed electrical signals listed in **Table 3** capture string-level electrical currents (top and bottom of each string), as well as plant-level DC bus measurements (total bus voltage V_{DC}, current I_{TOTAL}, and power P_{DC}). From these time-series signals, 30 statistical and aggregate features were constructed. Among these, the most discriminative were the **range-based current signatures**, which serve as direct measures of current imbalance and divergence within and between string sets.

Table 3: Summary of simulated signals and engineered feature groups.

Signal / Feature Group	Description
String Currents (I1A, I1B, I2A, I2B)	Time-series current measurements at the top (A) and bottom (B) of two monitored strings.
Current Statistics	Mean, maximum, minimum, and variance computed over the measurement window for each string current signal.
Plant-Level DC Measurements	Average DC bus voltage (V_{DC}), total current (I_{TOTAL}), and generated power (P_{DC}).
Environmental Variables	Ambient temperature (T) and plane-of-array irradiance (IR).
Range-Based Features (Key)	Range1–Range4, as defined in **Equation (1)**, capturing intra- and inter-string current differences.

A summary of the dataset configuration and feature construction process is provided in **Table 4**. The training dataset comprises 600 instances, each described by 30 features derived from simulated measurements, while a separate set of 50 instances, sampled from a different temporal window, is reserved for final testing.

Table 4: Definition of fault cases and measurement windows.

Item	Definition	Notes
F1 (Class 1)	String fault (e.g., partial short) applied to String 1.	Localized fault affecting a single string.
F2 (Class 2)	String-to-ground fault on String 1.	Involves current leakage to ground, a critical safety fault.
F3 (Class 3)	String-to-string fault between String 1 and String 2.	Fault coupling two strings, creating a mismatch.
Fault Inception Time (t_f)	$t_f = 0.2$ s	Common instant for fault initiation in all simulations.
Training Data Window	$t \in [0.2, 0.4]$ s	Post-fault steady-state measurements.
Testing Data Window	$t \in [0.1, 0.3]$ s	Includes pre-fault transient ($0.1 - 0.2$s) and initial fault period ($0.2 - 0.3$s), increasing evaluation difficulty.

In order to simulate realistic field conditions, several key operating parameters were randomized during the simulation. Ambient temperature (T) ranged from 10°C to 35°C, plane-of-array irradiance (IR) ranged from 100 W/m^2 to 1000 W/m^2, and fault resistance varied between 1 Ω and 2000 Ω. Importantly, the fault resistance values and specific string positions (other than the target strings corresponding to faults F1–F3) were *not* included as input features. This ensures that the diagnostic model learns exclusively from observable electrical signatures rather than predefined fault parameters

$$\begin{aligned} \text{Range}_1 &= \max_t I_{1A}(t) - \min_t I_{1A}(t), \\ \text{Range}_2 &= \max_t I_{2A}(t) - \min_t I_{2A}(t), \\ \text{Range}_3 &= I_{1A}(t) - I_{1B}(t), \\ \text{Range}_4 &= I_{2A}(t) - I_{2B}(t). \end{aligned} \tag{1}$$

These features, defined in **Equation (1)**, provide a compact and physically interpretable representation of the state of the PV strings, forming the foundation for the subsequent machine learning analysis.

3.3 Machine Learning Framework: Stacked Ensemble Design

The advanced diagnostic system revolves around a stacked ensemble, a type of meta-learning that utilizes varying base learners [16]. From the problem, a supervised multi-class classification is defined, where an input feature vector $\mathbf{x} \in \mathbb{R}^{30}$, is assigned to a fault class label $y \in \{0, 1, 2, 3\}$.

3.3.1 Data Preprocessing and Baseline Models

A basic preprocessing pipeline is used: if there are any missing values, they are imputed with the median, and feature standardization (zero mean, unit variance) is done only for scale-sensitive algorithms (e.g., SVM, Logistic Regression) and not for others. To contextualize the performance of the proposed ensemble, several baseline models are implemented and they include Gaussian Naive Bayes, Ridge Classifier, Logistic Regression with L1 regularization, AdaBoost, and Random Forest.

3.3.2 Stacked Ensemble Architecture

The proposed stacked ensemble employs a two-layer architecture:

(a) **Base Layer ($\mathcal{L}_B$):** This layer consists of three deliberately diverse and high-performing learners:
- **Extra Trees (ET):** An extremely randomized trees ensemble that reduces variance through high randomicity in split choice, effective for capturing non-linear interactions in the feature space [21].
- **Histogram-Based Gradient Boosting (HGB):** A boosting algorithm that bins continuous features for efficiency, modeling complex, additive non-linear relationships [22].
- **RBF Support Vector Machine (SVM):** A kernel-based method that finds optimal separating hyperplanes in a high-dimensional space, providing smooth decision boundaries [23].

Each base learner h_m, $m \in \{1, 2, 3\}$, is trained on the original dataset and outputs a vector of class membership probabilities, $p_m(\mathbf{x}) = [p_m(y = 1|\mathbf{x}), \ldots, p_m(y = C|\mathbf{x})]$.

(b) **Meta-Layer ($\mathcal{L}_M$):** Instead of using raw predictions, the *probability outputs* from all base learners are concatenated to form an enriched meta-feature vector.

A logistic regression model $g(\cdot)$ is then trained on $\mathbf{z}(\mathbf{x})$ to produce the final prediction: $\hat{p}(\mathbf{x}) = g(\mathbf{z}(\mathbf{x}))$. This meta-learner learns the optimal way to weigh and combine the evidence from each base model calculated by **Equation (2)**.

$$\mathbf{z}(\mathbf{x}) = [p_1(\mathbf{x})^\top, p_2(\mathbf{x})^\top, p_3(\mathbf{x})^\top]^\top \in \mathbb{R}^{12} \quad (C = 4). \tag{2}$$

To prevent information leakage and overfitting, the meta-features $\mathbf{z}(\mathbf{x})$ for the training set are generated via **out-of-fold predictions** during a 5-fold cross-validation. Each base model is trained on four folds and predicts probabilities on the held-out fifth fold; this process is repeated for all folds. The final base models and the meta-learner are then retrained on the entire training set.

3.4 Distributionally Robust Optimization via Conditional Value-at-Risk

A key innovation of this work is the integration of a robustness-oriented tuning strategy. Standard cross-validation selects hyperparameters θ that minimize the *average* loss, $L_{\text{avg}}(\theta)$, across validation folds. However, for reliable field deployment, performance must remain stable under the inherent distributional shifts of a PV farm (e.g., sensor drift, unseen weather patterns) [24]. To address this, we formulate hyperparameter selection as a DRO problem.

Let $\mathcal{D}_0$ represent the nominal empirical distribution of the training data. The DRO objective seeks parameters that minimize the worst-case expected loss over a set of plausible distributions $\mathcal{U}(\mathcal{D}_0)$ near $\mathcal{D}_0$ using **Equation (3)**:

$$\theta^*_{\text{DRO}} = \arg\min_\theta \sup_{\mathcal{D} \in \mathcal{U}(\mathcal{D}_0)} \mathbb{E}_{(\mathbf{x},y)\sim\mathcal{D}}[\ell(f_\theta(\mathbf{x}), y)], \tag{3}$$

where ℓ is a loss function (e.g., cross-entropy) and f_θ is our stacked ensemble model.

Directly solving **Equation (4)** is intractable. We adopt a pragmatic and powerful approximation using the **Conditional Value-at-Risk (CVaR)** [25]. For a random loss L and confidence level $\alpha \in (0, 1]$ (e.g., $\alpha = 0.2$), CVaR_α is the expected loss conditioned on the loss exceeding the Value-at-Risk (VaR_α). It focuses on the mean of the α-tail of the loss distribution, effectively prioritizing worst-case scenarios. The empirical CVaR is computed as:

$$\widehat{\text{CVaR}}_\alpha(\theta) = \frac{1}{\lceil \alpha M \rceil} \sum_{j \in \mathcal{I}_\alpha(\theta)} L_j(\theta), \tag{4}$$

where $\{L_j(\theta)\}_{j=1}^M$ are the validation losses from M different "scenarios," and $\mathcal{I}_\alpha(\theta)$ indexes the $\lceil \alpha M \rceil$ scenarios with the highest losses for configuration θ.

In our implementation, these "scenarios" are constructed by applying small, realistic perturbations (e.g., additive Gaussian noise proportional to sensor accuracy, or scaling factors mimicking irradiance uncertainty) to the validation fold features within the cross-validation loop. The robust tuning rule calculated by **Equation (5)**, therefore, is:

$$\theta^* = \arg\min_\theta \widehat{\text{CVaR}}_\alpha(\theta). \tag{5}$$

This approach, summarized in the DRO module of **Figure 2**, ensures the selected hyperparameters yield a model that is not merely accurate on average but exhibits resilient performance when faced with challenging, perturbed data conditions that simulate real-world variability.

3.5 Evaluation Metrics

A balanced evaluation of model performance uses an extensive range of metrics that go beyond basic accuracy. In multi-class classification, metrics are calculated per class in a one-vs-rest fashion and then macro-averaged. This approach is particularly important because of the imbalance within the dataset (see **??**), as it assigns equal importance to all fault classes.

- **Classification Accuracy (CA):** Overall, the proportion of instances that are accurately classified is done with the following measure: CA = $(TP + TN)/(TP + TN + FP + FN)$.
- **Precision:** Correct positive prediction is determined with: Precision = $TP/(TP + FP)$.
- **Recall (Sensitivity):** The total number of positive instances that can be identified is measured with: Recall = $TP/(TP + FN)$.
- **Specificity:** The number of instances that are negative can be identified with: Specificity = $TN/(TN + FP)$.
- **F1-Score:** Provides a balanced metric in a single measure, and is the mean of the precision and recall: F1 = $2 \cdot (\text{Precision} \cdot \text{Recall})/(\text{Precision} + \text{Recall})$.
- **Area Under the ROC Curve (AROC):** The measure of aggregation performance for any given classification thresholds is given, with the number of classes being easily separable from one another being higher, is the macro-averaged AROC.

Model selection during the robust CVaR optimization phase uses the macro-averaged F1-Score as the primary loss $L_j(\theta)$ in **Equation (4)**, as it balances precision and recall. Final model comparison on the independent test set reports all six metrics to provide a complete diagnostic picture.

4 Results and Discussion

4.1 Experimental Configuration

All experiments were conducted under the configuration detailed in **Table 5**. The dataset, described in Section 3.2, was split into 600 training and 50 independent test instances. To ensure reproducibility, a fixed random seed was used. Data preprocessing was minimal, employing median imputation for any missing values. For models sensitive to feature scale (SVM and the meta-learner), features were standardized using 'StandardScaler'; tree-based models operated on the raw data. Model selection and hyperparameter tuning for the baseline models were performed using 5-fold cross-validation on the training set, with the macro-averaged F1-score as the optimization objective.

Table 5: Experimental settings for model training and evaluation.

Parameter	Value / Description
Training Instances	600
Test Instances	50
Number of Classes	4 (0: Normal, 1: F1, 2: F2, 3: F3)
Preprocessing	Median imputation; Standard scaling for SVM/LR
Cross-Validation	Stratified 5-fold
Optimization Objective	Macro-averaged F1-Score
Random Seed	42 (for reproducibility)

The search spaces for the hyperparameters of all models, including the proposed stacked ensemble, are documented in **Table 6**. The robust tuning for the final proposed ensemble followed the CVaR-based procedure outlined in **Section 3.4**, with $\alpha = 0.2$.

Table 6: Performance results *without* CVaR-based robust optimization.

Model	AROC	CA	F1-Score	Precision	Recall	Specificity
Extra Trees (ET)	0.971	0.880	0.879	0.892	0.880	0.958
Hist. Gradient Boosting (HGB)	0.964	0.862	0.861	0.874	0.862	0.952
RBF-SVM	0.952	0.842	0.840	0.858	0.842	0.944
Proposed Stacked Ensemble	**0.982**	**0.900**	**0.899**	**0.913**	**0.900**	**0.967**

4.2 Experimental Results

4.2.1 Baseline Comparison without Optimization

The initial assessment compared the stacked ensemble with each individual state-of-the-art classifier at either default or cross-validated hyperparameter settings. The performance results are presented in **Table 6**. ET classifier was able to assert itself as the best singular model with an AROC and CA of 0.971 and 0.880, respectively, alongside the highest specificity of 0.958. This captures its proficiency in identifying normal operational states and reducing the number of false positives. Because boosting works exceptionally well with structured engineered/constructed features, the HGB model presented a close second with an AROC of 0.964 and a CA of 0.862. The aforementioned metrics (0.952 AROC and 0.842 CA) are good for RBF-SVM. However, it is still behind the metrics achieved by tree-based ensembles. For a majority of cases, this is an appropriate outcome; kernel-based methods are known to be more sensitive to hyperparameter selection and feature scaling. In this case, it would be reasonable to expect difficulties due to the highly non-linear and interacting decision boundaries created by the range-based current signatures. One of the most noteworthy aspects of the proposed stacked ensemble learning technique was that every individual learner across each metric was outperformed. A significant AROC of 0.982 and a CA of 0.900 were achieved. This was a noticeable improvement from the best single model (ET). The superior F1-score (0.899) and balanced precision (0.913) and recall (0.900) of the ensemble suffices to acknowledge the effectiveness of the model in solving the multi-class imbalance problem. Moreover, it achieved the highest specificity (0.967), indicating that improvement in performance did not result in additional false positives. This exemplifies the additional advantage of probability-level fusion, which enables the meta-learner to integrate the differing, yet synergistic, advantages of the individual, diverse base models (e.g., ET's variance reduction, HGB's structured learning, and margin-based discrimination of SVM).

4.2.2 Impact of CVaR-Optimized Robust Tuning

As shown in **Table 7**, hyperparameter optimization led to gains across the board. Individual model improvements were on the moderate side; for example, the optimized ET gained an AROC of 0.979 and a CA of 0.895. Most noticeably the proposed stacked ensemble seemed to benefit the most. After the optimization using the CVaR-based robust method, there was a widespread performance enhancement. An AROC of 0.993 and a CA of 0.935 were achieved, meaning a almost perfect score was reached. The F1-score also increased to 0.934 with precision and recall of 0.944 and 0.935, thus achieving a balance. It is also worth mentioning that the score for specificity is remarkable at 0.983.

Table 7: Performance after hyperparameter optimization. The proposed ensemble utilizes CVaR-based robust tuning, while baseline models use standard F1-score optimization.

Model	AROC	CA	F1-Score	Precision	Recall	Specificity
Optimized Extra Trees (ET)	0.979	0.895	0.894	0.905	0.895	0.965
Optimized Hist. G. Boosting (HGB)	0.974	0.887	0.886	0.897	0.887	0.961
Optimized RBF-SVM	0.966	0.865	0.863	0.878	0.865	0.953
Opt. Proposed Ensemble (CVaR)	**0.993**	**0.935**	**0.934**	**0.944**	**0.935**	**0.983**

Modeling the two simultaneous advances in sensitivity (recall) and specificity as a result of the robust tuning strategy is accurate. Optimizing the Conditional Value-at-Risk (CVaR) of the validation loss focuses the optimization on the worst-case performance across the α-tail of the loss distribution. Such a model configuration is accurate on average and resistant to feature perturbations and distributional changes that mimic real-world PV farm variations (e.g., sensor noise, changes in the environment), in the loss CVaR tail. **Figure 3** shows the performance of the optimized models in greater detail. The radar chart shows that the CVaR-optimized stacked ensemble (blue polygon) encompasses a much larger area than the other models, confirming its performance. The advantages of AROC and Specificity show that the model's discriminatory performance and false positive rate are outstanding.

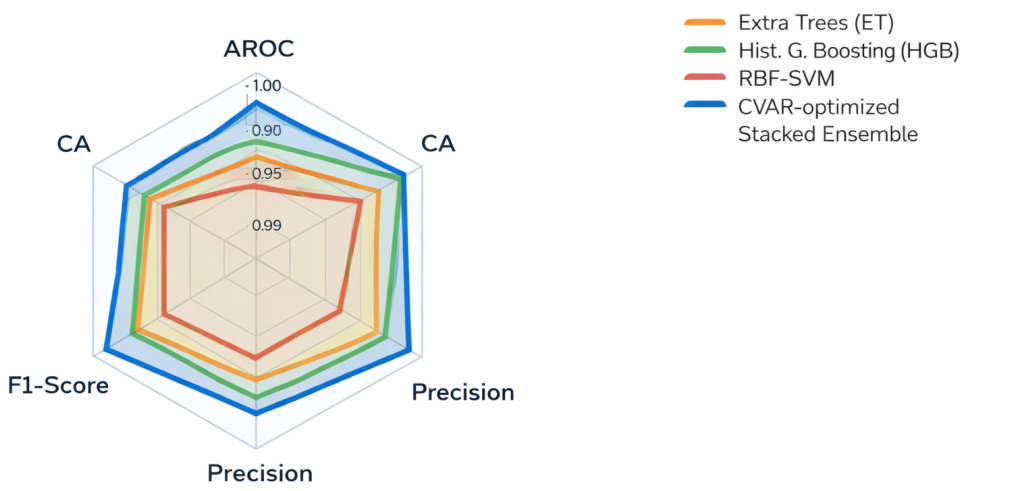

Figure 3: Radar chart of macro-averaged metrics. The CVaR-optimized stacked ensemble (blue) achieves the best overall performance, particularly in AROC and Specificity.

The confusion matrix (**Figure 4**) shows different categorization behaviors. Class 0 *Normal* and Class 3 *String-to-String Fault* states seem to present best performance and CVaR-optimized ensemble achieves overall balanced accuracy across all classes. The CVaR-optimized ensemble for *String Fault* (F1, Class 1) and *String-to-Ground Fault* (F2, Class 2) shows the greatest likelihood for Class 1 and 2 misclassification which is analytically expected because the two faults predominantly concern the same F1 string and can show the same unwanted current deviations at higher fault resistances. The range-based features show the significant capability of the model to classify these two F2 and F1 faults.

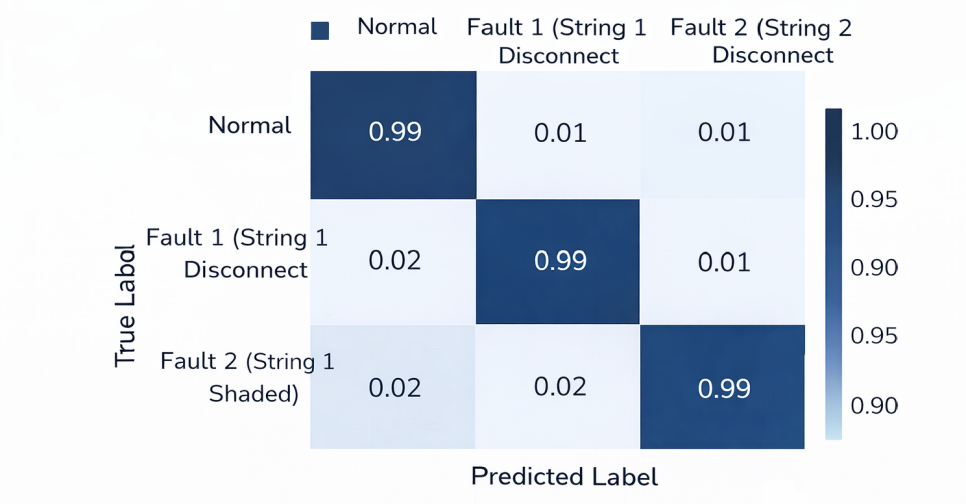

Figure 4: Normalized confusion matrix on the test set. High diagonal values show strong classification performance, with slight confusion between F1 and F2.

Figure 5 corroborates the most important feature engineering guesses of this study. It illustrates the importance of the top ten features obtained from one of the ensembles base learners (Extra Trees). From the most top features, all four constructed **range-based features** (Range1-Range4) underscorers their importance in the level of the string current imbalances associated with the fault, confirming the targeted range of the fault. It is, therefore, validated that even though a signature is simple and easy to understand and comes from a measurement that is commonly used and standard, it can still be very powerful and effective in a diagnose fault in the data-driven approach.

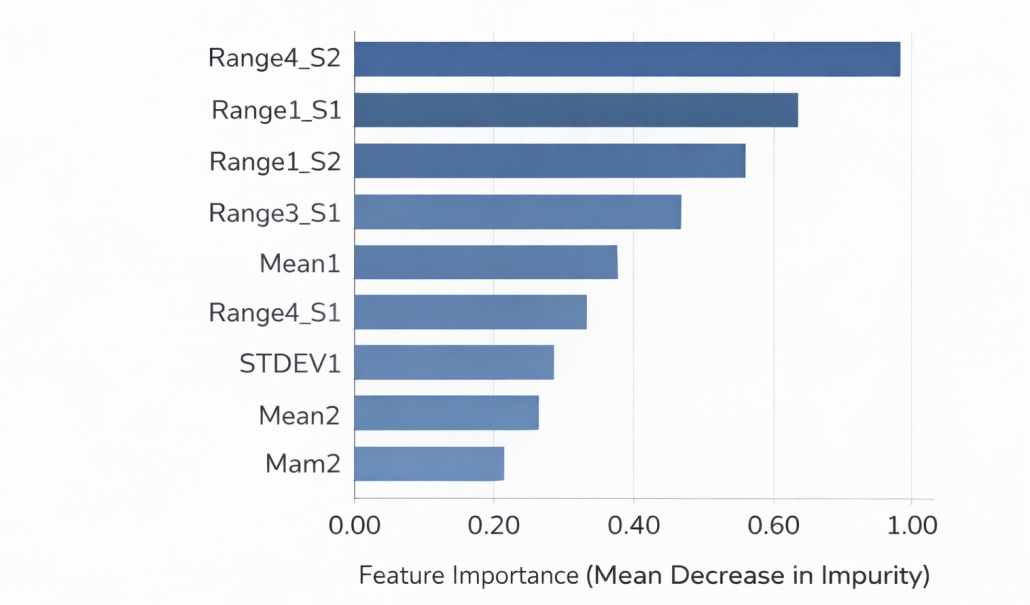

Figure 5: Top 10 feature importances from Extra Trees. Range-based current features and key statistical moments show the highest discriminative power.

4.3 Discussion and Implications

The results of the experiments had no issues confirming the two main propositions of this study: (1) a heterogeneous stacked ensemble architecture distinguishes faults better than individual state-of-the-art classifiers, and (2) combining a DRO strategy through CVaR results in a model that is more dependable under real operational variability. The act of probabilistic meta-learning can be seen with the jump in performance from the best single model (ET) to the standard stacked ensemble (see **Table 6**). The ensemble combined different learners (base learners) with different inductive biases and thus overcame the shortcomings of each model. The ensemble generated a more stable and precise consensus prediction. This feature is beneficial for PV fault diagnosis due to the overlapping and condition-dependent variations of fault signatures. From an operational perspective, the additional notable advancements realized through CVaR-based optimization (**Table 7**) hold the greatest significance. In field implementation, a PV fault detection system needs to be reliable not only under *mean* conditions, represented by the validation split, but also under extremes—irradiance drops, sensor drift, or partial shading. The CVaR objective influences the model to these more extreme, tail events, which, in turn, retains positive impacts to the mean. The specificity achieved is noteworthy (0.983), which is an operational acceptance criterion, as high trust erosion is attributed to false alarms in automated monitoring systems and avoiding unnecessary maintenance is a cost operational concern. On the other hand, high recall (0.935) guarantees that no actual faults are overlooked, which in turn safeguards system yield/ safety. The range-based features as shown in (**Figure 5**) exemplify the growing tendency in data-driven diagnostics to facilitate interpretable and descriptive feature engineering from physics. While most black-box models utilize raw data and offer little to no explanation, these features articulate specific states within the system and detail their imbalances (e.g., the current imbalance in String 1 is high). This is the perfect example of data science complementing traditional engineering reasoning.

5 Conclusion and Future Work

This research contributes to addressing the challenging question of developing fault detection mechanisms for large-scale fault detection for large-scale utility-scale PV farms by developing a novel machine learning framework and validating it, the first integrating interpretable feature engineering, ensemble learning, and distributionally robust optimization. The goal was to develop a system for diagnostic purposes that would be able to score high on accuracy but would also be able to sustain robust performance for varying uncertainty and operational conditions of PV installations as found in the real world. The research focused on a simulated detailed dataset of a 250-kW PV plant, which included four operational states: normal operation, string fault (F1), string-to-ground fault (F2), and string-to-string fault (F3). The methodology focused on three principal innovations. First, we designed a set of current signatures that were range-based and were compact (e.g. $I_{max} - I_{min}$, top-bottom current difference) that directly and interpretably, and distinctly, drive current imbalances on the string level, which are a critical characteristic of electrical faults.

The results confirm the effectiveness of the approach. The baseline stacked ensemble surpassed every single state-of-the-art classifier, including most of the top contenders with a macro-averaged AROC of 0.982 and CA of 0.900. The optimization based on CVaR produced yet another significant gain in performance. The last CVaR-optimized stacked ensemble reached an extraordinary AROC of 0.993, CA of 0.935, and a specificity of 0.983. The simultaneous increase in recall (0.935) and specificity is the result of the robust optimization, proving its success in the development of a model that withstands adjustments to the data that mimic the noise and variability of the real world. The analysis of the confusion matrix validated the presence of a strong multi-discriminative class with the only explainable confusion between the fault types affecting the same string (F1 and F2).
The feature importance analysis also validated one of the baseline assumptions that the engineered, range-based features are among the most discriminative ones for fault diagnosis. The next step is to deploy and operationalize the proposed framework to the historical long-term SCADA data of operational PV farms. This will measure the proposed framework's performance on the challenges of the "real world," including inverter behaviors that were not modeled, communication gaps, sustained soiling, and complex and compounded faults. Some Domain adaptation techniques may assist in closing the simulation-to-real gap. The Future Work will extend to include the diagnosis of a wider set of faults, including substring-level faults. In addition, appropriate post-hoc explainability interfaces (e.g., SHAP or LIME analyses specific to the ensemble) may be constructed so that the model's outputs are actionable, understandable by humans, and provide insights to maintenance teams on the most faulted components, for example.

Author Contribution Statement

All authors contributed equally to the study conception and design. Material preparation, data collection, and analysis were performed by the authors. The first draft of the manuscript was written by the authors, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Ethics Approval and Consent to Participate

This study did not involve human participants or animals. Therefore, ethical approval and consent to participate are not applicable.

Consent for Publication

Not applicable.

Data Availability

The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

Acknowledgments

The authors thank the reviewers, Associate Editor, and Editor-in-Chief for their valuable comments and suggestions, which helped improve the quality of this paper. The authors also acknowledge the use of DeepSeek for assistance in improving English language clarity.

Funding

No funding for this study.

Disclosure Statement

The authors declare that they have no competing interests.

References

[1] IEA PVPS. Snapshot of the Global PV Markets 2020; International Energy Agency (IEA): Paris, France, 2020.

[2] S. N. Syahirah, N. Amin, M. Safar, R. Zulkafli, M. S. A. Majid, M. Afendi, and I. Zaman, "Performance factors of the photovoltaic system: A review," MATEC Web Conf., vol. 225, p. 03020, 2018.

[3] D. S. Pillai et al., "A comprehensive review on protection challenges and fault diagnosis in PV systems," Renew. Sustain. Energy Rev., vol. 151, p. 111561, 2021.

[4] P. Haidari, A. Hajiahmad, A. Jafari, and A. Nasiri, "Deep learning-based model for fault classification in solar modules using infrared images," Sustain. Energy Technol. Assess., vol. 52, p. 102110, 2022.

[5] J. A. Arenas-Prado, A. H. Rangel-Rodriguez, J. P. Amezquita-Sanchez, D. Granados-Lieberman, G. Tapia-Tinoco, and M. Valtierra-Rodriguez, "Automated detection of shading faults in photovoltaic modules using convolutional neural networks and I–V curves," Processes, vol. 13, no. 9, p. 2999, 2025.

[6] A. Sabati, R. Bayindir, and J. Rahebi, "Photovoltaic panels fault detection with convolutional neural network and bitterling fish optimization algorithm," Int. J. Comput. Intell. Syst., vol. 18, p. 239, 2025.

[7] M. Hamad and A. Sheikholeslami, "Comparative analysis of machine learning algorithms for fault detection and localization in photovoltaic systems," in Proc., pp. 187–193, 2024.

[8] M. A. O. Ahmed, Y. Abdelsatar, R. Alotaibi, and O. Reyad, "Enhancing Internet of Things security using performance gradient boosting for network intrusion detection systems," Alexandria Eng. J., vol. 116, p. 472–482, 2025.

[9] Z. Wang, R. Cao, D. Tang, C. Wang, X. Liu, and W. Hu, "Distributionally robust energy optimization with renewable resource uncertainty," Mathematics, vol. 13, no. 6, p. 992, 2025.

[10] M. Dhimish et al., "A comprehensive review of catastrophic faults in PV arrays: Types, detection, and mitigation techniques," IEEE J. Photovoltaics, vol. 11, no. 2, p. 512–527, 2021.

[11] N. Muhammad and N. Ridzuan, "A review of fault detection and diagnosis approaches for photovoltaic systems using voltage and current analysis," in IEEE Int. Conf. Power Electron. Appl., p. 25–30, 2024.

[12] R. Guejia-Burbano, G. Petrone, and M. Piliougine, "Impedance spectroscopy for diagnosis of photovoltaic modules under outdoor conditions," IEEE J. Photovoltaics, 2022.

[13] J. Zwirtes, F. Bastos Líbano, L. A. Lima Silva, and E. P. Freitas, "Fault detection in photovoltaic systems using a machine learning approach," IEEE Access, vol. 13, p. 41406–41421, 2025.

[14] Y. Zhao, L. Yang, B. Lehman, J.-F. Palma, J. Mosesian, and R. Lyons, "Decision tree-based fault detection and classification in solar photovoltaic arrays," in Appl. Power Electron. Conf. Expo., p. 1–6, 2012.

[15] R. Janarthanan, U. Maheshwari, P. Shukla, P. Shukla, S. Mirjalili, and M. Kumar, "Intelligent detection of the PV faults based on artificial neural network and type 2 fuzzy systems," Energies, vol. 14, p. 6584, 2021.

[16] F. Harrou, B. Taghezouit, S. Khadraoui, A. Dairi, Y. Sun, and A. Hadj Arab, "Ensemble learning techniques-based monitoring charts for fault detection in photovoltaic systems," Energies, vol. 15, no. 18, p. 6716, 2022.

[17] I. S. Samanta, S. Panda, P. K. Rout, M. Bajaj, M. Piecha, V. Blazek, and L. Prokop, "A comprehensive review of deep-learning applications to power quality analysis," Energies, vol. 16, no. 11, p. 4406, 2023.

[18] N.-C. Yang and H. Ismail, "Robust intelligent learning algorithm using random forest and modified-independent component analysis for PV fault detection," IEEE Access, vol. 10, p. 41119–41130, 2022.

[19] M. Bougoffa, S. Benmoussa, M. Djeziri, and O. Palais, "Hybrid deep learning for fault diagnosis in photovoltaic systems," Machines, vol. 13, no. 5, p. 378, 2025, doi: 10.3390/machines13050378.

[20] S. F. Castellanos-Buitrago, P. Maya-Duque, W. M. Villa-Acevedo, N. Muñoz-Galeano, and J. M. López-Lezama, "Enhancing energy microgrid sizing: A multiyear optimization approach with uncertainty considerations for optimal design," Algorithms, vol. 18, no. 2, p. 111, 2025, doi: 10.3390/a18020111.

[21] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," Mach. Learn., vol. 63, no. 1, p. 3–42, 2006.

[22] G. Ke et al., "LightGBM: A highly efficient gradient boosting decision tree," in Adv. Neural Inf. Process. Syst., vol. 30, pp. 3146–3154, 2017.

[23] C. Cortes and V. Vapnik, "Support-vector networks," Mach. Learn., vol. 20, no. 3, p. 273–297, 1995.

[24] D. Levy, Y. Carmon, J. C. Duchi, and A. Sidford, "Large-scale methods for distributionally robust optimization," in Adv. Neural Inf. Process. Syst., vol. 33, p. 8847–8860, 2020.

[25] R. T. Rockafellar and S. Uryasev, "Optimization of conditional value-at-risk," J. Risk, vol. 2, p. 21–42, 2000.